Embedding novel and surprising elements in touch-screen games for children with autism: Creating experiences “worth communicating about”

Alyssa M. Alcorn

Doctor of Philosophy

Institute for Language, Cognition and Computation
School of Informatics
University of Edinburgh

2016

Abstract

Relative infrequency of communication initiation, particularly initiations that involve attention-sharing or other social purposes, appears to negatively impact the later-life outcomes of children with autism. Strategies to improve or encourage initiation skills in autism are hampered by the need for the behaviour to be spontaneous (i.e. unprompted by a partner). One potential approach that addresses the spontaneity issue is to extrinsically motivate initiations by changing aspects of the child's environment such that they merit, or even demand, initiating a communication. Detecting subjectively inconsistent (i.e. *discrepant*) aspects in game-like virtual contexts appears to be something that inherently interests young children with autism, and can motivate them to initiate spontaneous, positive communications. Initial evidence for discrepancy as a communicative motivator came from a study which re-analysed video data from an existing autism and technology project (ECHOES), illustrating that a heterogeneous group of children all reacted frequently and socially to naturally occurring (i.e. unintentional, non-designed) discrepant aspects within ECHOES. A set of high-level design principles was developed in order to capture "lessons learned" from ECHOES that might facilitate re-creation of a similar pattern of spontaneous, positive initiation around discrepancy. A second, proof-of-concept study implemented these design principles in a set of three new touch-screen games (*Andy's Garden*) that sought to establish, and then deliberately violate, child expectations (i.e. provide *discrepancy-detection opportunities*: DDOs). Children reacted socially and positively to the new games and DDOs. The results of this study allow us to answer its overall questions affirmatively: it is possible to motivate children's communication–specifically, their initiation–by including deliberately-designed DDOs in a set of games. These findings are the first step towards determining whether discrepancy-detection opportunities may form a component of a future technology-based communication skills intervention, capable of changing children's initiation behaviour outside of a game context.

Lay Summary

Autism is a condition that affects many aspects of a person's development and cognition, particularly social skills and communication. Starting a new communication with another person (i.e. *initiating communication*) is generally very difficult for people with autism. They may initiate communication less often compared to people without autism. Studies on children with autism suggest that difficulties with initiation can negatively impact them throughout their lives. It is a research priority to help children with autism learn or improve initiation skills. Researchers and educators have tried many strategies to help children with autism initiate communication. One strategy has been to change parts of a child's physical environment in ways that make him or her want or need to initiate communication. For example, setting up a situation where a daily routine is changed, or the child cannot do something herself and must ask for help.

In previous research, the ECHOES computer game allowed children to play with objects and a character (Andy) in a "magic garden". Video of children with autism interacting with ECHOES suggests that changes, surprises, new things, and "mistakes" within a computer game can also interest them and make them want to communicate. For example, some children exclaimed and corrected Andy when he unexpectedly tried to put a red ball into the yellow box during a sorting game. Young children with autism repeatedly initiated about these kinds of events to an adult researcher, or to Andy. These "mistakes" and changes in the ECHOES game were due to software problems, and were not a planned part of the design. Children initiating communication about these events was a very positive result, but unexpected. This raised the question of whether it was possible to purposely create these kinds of events within a game—and whether they would motivate children to initiate.

The first step was to develop recommendations, based on the ECHOES videos, about how a new game could try to re-create the same type of child communications. These recommendations were used to create three new, small games (*Andy's Garden*). Each game deliberately included changes, mistakes, and surprising events, similar to the ones that happened by accident in ECHOES. Ten children with autism were videoed playing the new games over several days, to determine whether they initiated communications about any of these events.

Overall, children found the new games enjoyable, even funny. They noticed and chose to communicate about many of the changes, surprises, new things, and mistakes deliberately included in *Andy's Garden*. Individual children varied in what parts of

the game interested them, and how much they initiated to the adult researcher versus the Andy character. According to these results, it is definitely possible to motivate children with autism to initiate communication by introducing mistakes and similar events into a computer game. They appear to find these things “worth communicating about” while they play. More work is needed to discover whether future games could use this strategy to teach children with autism *new* initiation skills, or help them start more communications in daily life, when they are not playing the game.

Acknowledgements

This thesis has one name on it, which obscures the extensive time and effort contributed by many people over the past several years. Grateful thanks are due to the patience, reassurance, and long experience of my supervisors, Prof. Helen Pain and Dr. Judith Good (who went far beyond the call of duty as a second supervisor). My teaching mentor, Dr. Stratis Viglas, gave regular doses of common sense and perspective, especially in the early stages. My colleagues Dr. Sue Fletcher-Watson and Dr. Christopher Frauenberger contributed advice and expertise. Mihaela Dragomir very generously spent many hours helping with Study 2, and Jasmin Mahmoodi worked on second coding of Study 1 data. Many further colleagues gave feedback, tested buggy software, and rendered other support during this work, particularly Mihaela Dragomir (again!), Catherine Crompton, Dr. Aurora Constantin, Dr. Robin Hill, and all the other LAER Lab members from Informatics and Education. Many colleagues in the IDC and AIED communities, in ASC-tech, and from the ECHOES project have also been an important sounding board and source of ideas over the past years. The following people gave important personal support at various stages, and also deserve thanks: AD, CC, ER, GM, HH, JG, SM, TJ. And the people at home who helped me get here: ALLT, AG, CT, RGN, TM, Mom, Dad, Grandpa Rex, and my early research mentors Dr. Lucy Troup (Colorado State) and Dr. Cedar Riener (Randolph Macon, formerly of Mills College). Further thanks are due to the enthusiastic children and parents in the Informatics Forum, especially to Leon Blake, the new voice of Andy. Finally, infinite thanks to the participating special schools, staff, and children with ASC, who gave generously of their time and energy during the original ECHOES and the *Andy's Garden* evaluation. This work is ultimately for you. This research was funded by the University of Edinburgh and the Scottish Informatics and Computer Science Alliance (SICSA). It includes materials from the ECHOES project (funded by ESRC/EPSRC TEL; RES-139-25-0395). The ECHOES environment was designed and developed by the ECHOES team, including the author (see www.echoes2.org), with contributions from autism and education community stakeholders.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(Alyssa M. Alcorn)

The following peer-reviewed publications during the period of PhD study have reported versions of the material included in this thesis. Their citations are listed along with the thesis author's contribution. Those labelled as the author having "main responsibility" means primary responsibility for the analysis and interpretation of the reported data, authorship of the majority of the final published text, and production of associated figures and tables.

Paper information	**Author's role in publication**
Alcorn, A.M., Pain, H., & Good, J. (2014). Motivating children's initiations with novelty and surprise: Initial design recommendations for autism. *In Proceedings of the International Conference on Interaction Design and Children* (IDC '14, Aarhus, Denmark) pp 225-228. DOI: 10.1145/2593968.2610458	Main responsibility for all aspects of this publication. No new data analysis.
Alcorn, A.M., Pain, H., Good, J. (2013). Discrepancies in a Virtual Learning Environment: Something ``Worth Communicating About" for Young Children with ASC? In *Proceedings of the International Conference on Interaction Design and Children* (IDC '13, New York) pp 56-65. DOI: 10.1145/2485760.2485783	Main responsibility for all aspects of this publication.
Alcorn, A.M. (2013). Deliberate Discrepancies as a Design Strategy for Motivating Social Communication in Virtual Environments for Young Children with Autism. Doctoral Consortium short paper. In *Proceedings of the International Conference on Interaction Design and Children*. (IDC '13, New York) pp 659-662. DOI: 10.1145/2485760.2485876	Main responsibility for all aspects of this publication.
Alcorn, A.M., Good, J., & Pain, H. (2013). Deliberate System-Side Errors as a Potential Pedagogic Strategy for Exploratory Virtual Learning Environments. In H. C. Lane et al. (Eds.): *AIED 2013, Lecture Notes in Computer Science*, Volume 7926, pp 483-49. DOI: 10.1007/978-3-642-39112-5_49	Main responsibility for all aspects of this publication.
Alcorn, A.M. (2013). Discrepancy-detection in Virtual Learning Environments for Young Children with ASC. Doctoral Consortium short paper. In H. C. Lane et al. (Eds.): *AIED 2013, Lecture Notes in Computer Science*, Volume 7926, pp 884-887. DOI: 10.1007/978-3-642-39112-5_137	Main responsibility for all aspects of this publication.

For Little Grandma, who would have been amused by *Andy's Garden* and would have approved of my grand adventure in Scotland.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Unexpected observations in the ECHOES virtual environment

A young child is playing an activity on a large touch screen, apparently highly engaged and having a good time. A 3D character (Andy) is helping him to sort coloured balls into boxes, but is not doing a very good job—he is about to put a blue ball in the yellow box. In the background, a researcher looks on completely horrified, as though she wants stop what is about to happen. What is going on here?

The child playing the activity has an *autism spectrum condition* (ASC)[1]. These are pervasive developmental conditions, characterised by notable difficulties in communication and social interaction, plus the presence of repetitive behaviours and/or interests. A core characteristic of autism is an "insistence on sameness [and] inflexible adherence to routines", sometimes including "extreme distress at small changes" (APA, 2013, p.50). With this information in mind, we would expect a child with

[1]According to recent survey research, "there is no single way of describing autism that is universally accepted and preferred by the UK's autism community" at this point in time (Kenny et al., 2016, abstract). This thesis uses the term *ASC* rather than referring to *autism spectrum disorder* (ASD). Both terms are widely used in autism research and practice literature. Here, *autism spectrum conditions* is seen as preferable because the word *disorder* conveys a very negative, medical view of autism as a disease to be treated or cured. *Condition* is more neutral, and still makes clear reference to the concept as it is defined in the literature. With a few exceptions, the thesis also refers to *children with autism*, rather than *autistic children*. This "person-first" language has been chosen with the intent to respectfully acknowledge individuals as individuals, rather than only as exemplars of autism. It is worth noting that person-first language is in itself contested: according to another view, autism (and other disabilities) are identities and ways of being, and speaking of a "person *with* autism" is a false separation. Disability-first language such ("autistic person") would be preferable. Kenny et al. (2016) note that there appears to be a split, in which researchers and practitioners may be more likely to follow the language espoused here (ASC, person with autism), whereas adults and families on the autism spectrum may support disability-first language.

Figure 1.1: A child points out the correct box following Andy's ball-sorting mistake. This example is discussed in more detail in Chapter 4.

ASC to be very unhappy when Andy makes a sorting mistake. He has done it right every other time! We would expect that the system changing and breaking its own rules, apparently at random, would be a stressful event for him. The researcher in the background is anticipating that a negative reaction might be imminent. What actually happened was far more unexpected: the child found Andy's mistake hilarious, laughing loudly and pointing to the box he should have used (Figure 1.1). Moreover, this was not an isolated example of children with ASC reacting positively to events of this type.

During the evaluation of ECHOES *virtual environment* (VE) for children with autism, researchers noted that the behaviour of both the VE and *virtual character* (VC) were repeatedly but unpredictably altered due to errors in the software. Examples of the latter are that the virtual character almost always demonstrated the correct activity actions but sometimes made "mistakes" (as in the opening example), or that objects might be touch-unresponsive or mysteriously absent from the screen. There were multiple examples of children reacting to these situations, which were often strikingly inconsistent with the body of knowledge they had developed about the VE's routines, patterns, and object properties. Frequently, the reactions were spontaneous initiations to the researcher(s) and/or to the VC. For example, Ethan[2] initiated non-verbally to the researcher when the character Andy appeared to be missing at the start of an activity. After watching the screen for almost 10 seconds, waiting for Andy to enter and introduce the activity[3] Ethan finally turned all the way round to look at the researcher, seated behind him. He continued looking until she responded to him, approaching and leaning down to tell him "Wait 'til Andy comes back."

Children also showed particular interest in communicating about *novel* behaviours and objects that were introduced in different ECHOES activities, as a deliberate part of the design. A good example of a child initiating to someone about a novel element is Anthony's reactions to the fireworks reward. In the ball-sorting game mentioned in the opening example, a child would get an attractive reward when all the balls of one colour were put away. The blue box became an animated fireworks display with sound effects. The first time Anthony got this reward, he immediately shouted "BOOM!" (imitating the sound effect) and turned to look at the researcher, exclaiming "fireworks!" and clapping his hands. He immediately turned to initiate again to the second researcher behind the camera (Figure 1.2), sharing his obvious excitement. He explains to both researchers, "is cause the box go BOOM!" and throws his arms out in an exploding gesture. Here, this event has been *so* interesting to Anthony that he initiates multiple times, using several different behaviours.

Not all child reactions to unexpected or novel aspects were directed to others. Sometimes children had clear but non-social reactions. For example, Lucy was excited to discover that she could move flower pots in the stacking game by "swiping" at them with big, physical movements. Focused completely on the screen, she loudly says "wow!" and tries the action over and over (Figure 1.3). Here, she clearly reacts to this new effect, but does not share the experience with another person.

Observations of instances like the ones described here are notable for two main reasons: First, because they run contrary to the widely accepted and documented "insistence on sameness", which is included in the criteria for autism diagnosis, *restricted and repetitive behaviours* (APA, 2013, p.50). Furthermore, research has shown that predictable sequences of events in the external environment lead to increased responsiveness of children with autism, as compared to unpredictable sequences (Ferrara and Hill, 1980). As noted at the start of the chapter, autism literature might lead us to

[2]All child names used in this thesis are pseudonyms. All children pictured in the thesis have parent consent for use of photos for research and teaching purposes, including publications.

[3]Andy should enter five seconds after the activity start, and this is a noticeable deviation from the pattern established in previous activities.

Figure 1.2: After initiating to the researcher at the computer about the fireworks reward, Anthony turns to share his excitement and positive affect with the researcher behind the camera.

Figure 1.3: Lucy discovers that she can move flower pots with a swiping gesture, and exclaims “wow!”

expect that children would be upset and uncomfortable about encountering novel or expectation-violating aspects of a game that had previously behaved in a predictable and lawful way. Overall, these occurrences within the ECHOES dataset appeared to be perceived as fun or humorous, rather than disruptive.

The second reason that these observations are unusual is *how* children have reacted. In many cases, they were spontaneously initiating to social partners[4] about what was happening—or not happening—in the environment. Initiation of new communications, rather than responding to a partner's communication, is particularly difficult for this group. Children with ASC tend to initiate all types of communication infrequently compared with *typically developing* (TD) or developmentally delayed peers (Mundy et al., 1990). Based on their reactions, situations with novel or expectation-violating aspects appear to be something "worth communicating about" for children with autism. These opportunistic observations suggest that the opportunity to discover when "something is amiss" may be inherently interesting for this group. This is a potentially valuable phenomenon: social communication can be considered a foundation upon which more advanced skills are built. There is widespread agreement that early childhood interventions to improve social communication, including initiation, provide a better chance of positive outcomes later in life (e.g. Parsons et al., 2009a). Learning about a new way to motivate children to use their initiation skills, especially if the initiations are for social purposes, could be a valuable contribution to future technology designs and (potentially) intervention.

These novel or expectation-violating game elements to which children reacted in ECHOES are instances of *discrepancy*[5]. Novelty and expectation-violation are both subjective: they are determined in relation to what a particular individual knows and expects. Discrepancies are situations in which a particular child considers some aspect of the VE to be *subjectively inconsistent* with his current knowledge and experience—either because it violates expectations or is not yet included in his expectations (i.e. is novel). This emphasis on subjectivity is important, as it means that discrepancies are *states of affairs* that may occur while a particular child is interacting with a particular environment, not objective "things" that exist in the world. Our information about discrepancy is indirect, through a child's reaction. Thus, the main unit of analysis is the *discrepancy- child reaction pair* (DR pair), not discrepancy alone.

[4]This term is defined in Chapter 4; see section 4.3.4.

[5]The discrepancy concept was developed during the course of the current work (see Chapter 4), but is previewed here to help readers understand the investigation proposed in the research questions and introduce key terminology.

A main goal of the current work has been to systematically define and investigate the phenomenon (discrepancy) represented by these initial observations in ECHOES, and determine the extent to which it appears to motivate children's communication.

1.2 Research Questions

The concept of discrepancy, and particularly discrepancy as a way to motivate communication, was developed and tested in the course of the current work. Rather than being planned as a whole, the work has been conducted as a series of stages, with each stage posing questions and developing working hypotheses to guide subsequent stages. The research questions capture the chronological investigation followed throughout the thesis, with different chapters or sets of chapters addressing each issue. They are a *reconstructed logic* of the current research, based on its final form, rather than a *logic-in-use* (Kaplan, 1964), though the individual chapters reveal much more of the latter reasoning.

Following from the initial discrepancy-reaction observations in the ECHOES data:

RQ1 What is happening when children with ASC were observed to communicate about novel and apparently surprising aspects of ECHOES? What is the nature of this phenomenon (eventually labelled as discrepancy)?

RQ2 Is discrepancy represented by a few rare but striking instances in the ECHOES dataset, or are there sufficiently widespread examples to merit further investigation and form a basis for categorisation?

RQ3 Given that instances of discrepancy appear repeatedly across ECHOES sessions and participants, how can we describe and categorise them? Which features or sources of information need to be included in order to capture the nature of these instances?

RQ4 Is it possible to create discrepancy and discrepancy-contingent communications on purpose? How might this be achieved?

RQ5 Guided by design principles rooted in empirical work with ECHOES and autism theory, can designed *discrepancy-detection opportunities* (DDOs) successfully create a *motivating but manageable* experience with new participants, in a new set of games?

As should be clear from the research questions, the current work is about defining and systematically exploring the new concept of discrepancy. It does not aim to create an intervention to *change* child behaviours, but instead focuses closely on child behaviour within a game context, and how certain elements of the design may motivate communication. The issue of intervention is reserved to future work. Chapter 3 returns to these questions, and identifies key sub-questions for each.

1.3 Thesis outline

The remainder of this thesis investigates these questions in order, and provides further explanation of how each question helped to frame the subsequent research questions. Chapter 2 reviews relevant background literature on the characteristics of the autism spectrum, focusing on issues around social communication. It also considers technologies for autism, and the existing work that has tried to motivate children's communication by modifying their environments. Chapter 3 describes the overall research approach taken in the thesis, placing it in context of the autism and technology subfield. Chapter 4 begins the main body of the work. It returns to the ECHOES dataset and observations mentioned in this chapter and describes how they were used as the basis for iteratively developing the "core concept" of discrepancy, the taxonomies of discrepancy and child reaction, and an annotation scheme for identifying and labelling these instances in video data. Chapter 5 (Study 1) reports the results of systematically applying the final annotation scheme to a subset of ECHOES data, with a focus on children's communication. Chapter 6 moves into the design-focused part of the project, presenting additional analyses of specific discrepancies in ECHOES, and reflecting on the apparent contradiction between the "need for sameness" and children's enjoyment of unexpected and novel events in ECHOES. To support transfer of positive, ECHOES-like patterns of interaction to new contexts, the chapter presents six high-level design principles. The principles are implemented in a new set of games with discrepancy-detection opportunities, described in Chapter 7. Chapter 8 reports the games' evaluation in a proof-of-concept scale school study (Study 2), including results for child communication, and more design-oriented results that look at the success of specific design choices. Finally, Chapter 9 revisits the line of investigation summarised by the research questions, identifying key findings for each one. There is general discussion of the current findings and their implication for ASC-tech and autism research more generally. A future work section describes additional analyses of current data, and longer-term investigations that could explore discrepancy as an intervention strategy.

Chapter 2

Reviewing the intersection of autism, communication, and technologies

2.1 Introduction

This chapter reviews relevant background literature to place the thesis work in context. Some additional literature is introduced in later chapters, where it is most pertinent to the work being conducted. As the current investigation is interdisciplinary, this review presents work in several different areas (general autism background, social communication interventions, technology for autism [ASC-tech]) and then includes a short discussion to help synthesise this diverse information. In every area, the information presented narrows down quickly to the current focus: spontaneous initiation of communication. It does not address broader socio-communicative issues in autism outside of initiation, and presents the argument for applying technologies to ASC, supported with several focused examples of ASC-tech for communication.

The one area that is discussed in depth is existing work on environmental modification, or, making changes to a child's familiar environment to try to motivate or even "demand" communication in some way. The idea of using environmental factors to motivate communication surfaces in a number of authors' work over more than thirty years, but the picture is fragmented, confusing, and sometimes repetitive. It is untangled here as far as is possible, because this work is the closest precedent in the autism literature to the concept of discrepancy that has been introduced briefly in Chapter 1.

2.2 Overview of autism spectrum conditions (ASC)

2.2.1 Core characteristics and diagnosis

The autism spectrum is a group of pervasive developmental conditions, all of which are characterised by notable difficulties in communication and social interaction, plus the presence of repetitive behaviours and interests. Due to the heterogeneous nature of the autism spectrum population, two individuals with the same overall diagnosis may have little in common otherwise in terms of both their strengths and qualitative impairments. The current DSM-5 diagnostic criteria for the autism spectrum conditions are behavioural, and divided into two main categories[1]. First, "persistent deficits in social communication and social interaction across multiple contexts", which may include difficulties related to emotional understanding and expression, relationships (of any kind), non-verbal communicative behaviours (e.g. gesture, eye contact), and initiation or response to interactions (APA, 2013, p.50). As the current work is concerned fairly narrowly with children's initiation of communication (both verbally and non-verbally), the literature review does not go into depth regarding other types of social communication and interaction issues (e.g. peer relationships, eye contact, emotion recognition). Initiation in autism is discussed in Section 2.3.

The second diagnostic criterion is "restricted, repetitive patterns of behaviour, interests, or activities" (p.50). For diagnosis, an individual must show at least two types of behaviour from a list that includes repetitive or stereotyped actions or speech, abnormally focused and restricted interests, and "insistence on sameness" (p.50). This last manifestation is central to the current work, and is discussed in more detail below. For both categories of diagnostic criteria, the DSM-5 stresses that these deficits must cause "clinically significant impairment" to an individual's functioning in one or more areas (APA, 2013, p.50). Many widely-used behavioural assessment tools to diagnose the autism spectrum conditions focus on gathering information about these categories, particularly how social behaviours are—or are not—used. Examples include the Autism Diagnostic Interview–Revised (ADI-R; Rutter et al., 2003; Lord et al., 1994) and the Social Communication Questionnaire (SCQ; Rutter et al., 2003), both based on parental reports, and the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000), which involves structured child observation by a clinician or other

[1] In the previous DSM-IV (DSM-IV, 1994), there were three main categories: *qualitative impairment in social interaction*, *qualitative impairment in communication,* and *restricted repetitive and stereotyped patterns of behaviour, interests, and activities*. These criteria were used until 2013.

professional.

While not part of the diagnostic criterion, two more issues should be mentioned as part of an overview of ASC: intellectual disability and language difficulties. Both are extremely widespread in ASC and strongly influence an individual's "ability" in different contexts. The DSM-5 (2013) advises that a diagnosis of Autism Spectrum Disorder should specify whether it is "with or without accompanying language impairment" and "with or without accompanying intellectual impairment". Intellectual disability, which is "characterised by deficits in general mental abilities" (APA, 2013, p.31) affects a significant number of individuals on the spectrum; Levy et al. (2009) estimates as many as 40-80%. Due to this prevalence, the concept of *developmental age* or *verbal-mental age* can be very useful, and give a better estimate of a person's abilities than does chronological age.

Levy et al. (2009) give an estimate of 50-63% of people with ASC being affected by some kind of language deficit[2]. Tager-Flusberg, Paul, and Lord (2005, as cited in Eigsti et al., 2011) suggest about 25% of people with ASC do not acquire verbal language at all. Verbal language acquisition may be atypical or substantially delayed. Language use is also frequently limited and atypical, compared to usual development. It may be stilted, repetitive, and overly literal. Metaphors and humour may both be quite difficult or simply not understood. Turn-taking and repair in conversation may also be notably impaired. For a brief overview of this literature and a specific review of language acquisition in ASC, see Eigsti et al. (2011).

It is necessary to keep in mind that there may be a split between children's language comprehension (i.e. receptive language) and the language they are able to produce. The DSM-5 advises that these should be considered separately. Perhaps surprisingly, it is often *receptive* language that is relatively delayed in autism. A child may speak relatively fluently, but have a much less complete understanding of language. In some cases, the language children produce may be *echolalia*, or imitation of language they have previously heard elsewhere (e.g. Schuler, 1979; Prizant et al., 2005). This may be immediate or delayed, exact or somewhat modified. It is important to note that echolalia frequently *is* used in a communicative way, even if this is highly idiosyncratic and not readily understood by others people (Prizant and Rydell, 1984; Prizant and Duchan, 1981).

The practical implication of these language difficulties is that, when working with children with ASC, language is most likely to be understood when it is kept simple in

[2] Based on the DSM-IV (1994). These prevalences may now have shifted.

terms of grammar and vocabulary. It should be direct, giving instructions or making specific requests, rather than relying on implicature. Even those individuals considered "high functioning" may show atypical language use and understanding. Language is an important consideration in relation to designing games, and interpreting children's behaviour as they interact with the games and the researcher.

2.2.2 Restricted and repetitive behaviours (RRB) and sameness

The diagnostic criterion that is most relevant for understanding why enjoyment of discrepancy is unusual for the ECHOES participants would be the second DSM-5 category *restricted and repetitive behaviours*, sometimes abbreviated RRB. It has several aspects, including what are often called "special interests", and sensory issues. The category encompasses an extremely wide range of behaviours that may not initially appear to have a close relationship to one another. Repetitive, self-stimulatory behaviours like rocking or waving fingers in front of the eyes are a part of this category, but so too are the vivid anecdotal examples from the literature, that refer to individuals memorising decades worth of train schedules, or only eating orange-coloured foods, or carrying a small vacuum cleaner at all times (Schreibman, 2005). Any general resource on autism includes similar examples (e.g. Schreibman, 2005; Howlin, 1998; Wall, 2010). The following discussion focuses on a specific aspect of RRB, the "need for sameness" or, in Kanner's famous description, "an *anxiously obsessive desire for the maintenance of sameness* that nobody but the child himself may disrupt on rare occasions." (Kanner, 1943, p. 245, emphasis in original). The DSM-5 (APA, 2013) explains this aspect of RRB as follows:

> "Insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or non-verbal behaviour (e.g. extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take the same route or eat the same food every day)" (p. 50, criterion B.2).

Therese Jolliffe, an adult academic with autism, offers some insight into the importance of these behaviours as an important coping mechanism in a frightening and disordered world. She explains that "reality to an autistic person is a confusing, interacting mass of events, people, places, sounds, and sights. There seems to be no clear boundaries, order, or meaning to anything....Set routines, times, particular routes and rituals all help to get order into an unbearably chaotic life. Trying to keep everything the same reduces some of the terrible fear" (Jolliffe et al., 2001, p. 50; quote reproduced from Jolliffe et al., 1992). Consequently, the current consensus in the autism research community is that it is counter-productive, even cruel, to try to "extinguish" routines and rituals without guiding a person to a more appropriate alternative.

The phrase "insistence on sameness" can obscure the enormous—often cumulative—effect that routines, non-functional rituals, and change aversion can have on a child with ASC, and the people around them. While individual preoccupations, rituals, or aversions to change may be limited in scope, a child may have many of them, which can add up to a very difficult situation in the home or the classroom. In a family where someone with autism may be obsessed with maintaining sameness in the house and may experience intense anxiety at any change, the family may effectively be prevented from replacing old or broken housewares, re-painting, or moving furniture for years at a time. Some routines or rituals can seem fairly harmless but can build up to disaster, such as when the *only* item of clothing the child will wear finally falls to pieces, or a favourite TV show is cancelled. A more verbal child might show these repetitive and stereotyped behaviours by restricting his or her speech and insisting that others do the same, always using certain phrases or giving certain responses. Howlin (1998) notes that repetitive questioning (e.g. about certain events or topics) can also be a very common verbal routine. These behaviours can be a source of significant concern and stress to parents and teachers of children with autism, not least because those children will grow up. Behaviours that may amuse strangers when exhibited by a toddler may be considered threatening when they are exhibited by an adult.

In order to try to reduce the impact of these behaviours on the individual's life (and often those of their families!) many resources for parents and practitioners emphasise the importance of predictability, routines for daily life (e.g. getting ready for school), and clearly-signposted, supported transitions. Many practice-focused sources also extensively discuss how to introduce gradual and manageable changes, such as moving a child from accepting a single and *exact* time or location for things (e.g. for a meal to start) to accepting a narrow range of alternatives, and then a larger range (e.g. Wall, 2010; Howlin, 1998). Howlin (1998) also points out that environmental changes (e.g. moving to a new classroom) can sometimes be a good opening for creating behavioural changes. As many routines may be limited to certain situations or locations, they may not transfer elsewhere. Alternately, the child may be guided into establishing a new (and hopefully less disruptive) routine if prevented from settling in to a familiar pattern.

Given the important role of rigid, stereotyped behaviours and routines in creating structure and reducing anxiety, the obvious enjoyment of surprising, routine-violating events observed in ECHOES and *Andy's Garden* is puzzling. The literature would, most likely, lead us to predict that some of the apparently random variations would be a source of discomfort. We might anticipate something like Kanner's description of children playing with blocks, in which "the absence of a block or the presence of a supernumerary block was noticed immediately... If someone removed a block, the child struggled to get it back, going into a panic tantrum until he regained it" (Kanner, 1943, p.245). In the data so far, this is not the case. Chapter 6 returns to this question in detail, formulating a working hypothesis that, under the right circumstances, change and difference can interest and motivate children when they are counterbalanced by substantial sameness in the environment.

2.3 Initiation of communication

2.3.1 Contingency, spontaneity, and initiation

In order to understand what spontaneous initiation means in the current research context, it is useful to consider two dimensions along which communicative behaviour may vary: spontaneity–promptedness and initiation–response. Here, these are treated as binary distinctions because that is sufficient for the current analyses, though finer gradations are possible[3]. Both parameters assume that the behaviour being discussed is *intentional*, in Tomasello's sense of being intended and purposeful (1995), and that it is also *socially directed*, which is to say that it is directed to a partner[4] and not to the self, and intended to convey something. Both parameters tap into the same underlying idea, which is that of *contingency*[5] and the relationships between interactions.

For any behaviour to be *prompted* means that it has been cued in some way. That behaviour has been preceded by or co-occurs with a stimulus that attempts to guide it toward some particular result that is considered relevant or correct (Cooper, 1987, as cited in Chiang and Carter, 2008). A prompt could refer either to the stimulus itself, or to the act of providing this stimulus to the child. While prompts may be inanimate and may exist in the environment, such as a school sign that reminds children about

[3] For example, see Chiang and Carter (2008) on non-dual conceptualisations of spontaneity.

[4] This term, used interchangeably with *social partner* is defined in Chapter 4; see section 4.3.4.

[5] Some senses of the word contingency have to do with accident and chance. Here, it is meant in the sense of some event being dependent upon previous events or circumstances.

queueing behaviour, the prompts that are of interest in this section are those which are actively being delivered by a (usually adult) social partner, to the child.

A clear example of prompted communication would be a parent telling the child "Say thank you", and the child saying "Thank you". Responding to a question could also be considered prompted communication, as it also attempts to elicit a certain type of response. In the literature on interventions and teaching programmes for people on the autism spectrum, *prompt* is often used to mean these types of directive cues for specific behaviours. The use of prompts in their broader sense is widespread within autism research and practice, but often they often have special vocabulary (e.g. a *visual support*). While a verbal response in ordinary conversation is prompted in the sense that it is contingent on previous conversational "moves", it would be odd (at least in the autism literature) to say that it was prompted. References to prompting later in this document or in the course of this project mean the term in its most overt, directive sense. To avoid any confusion, the majority of the discussion of initiations and responses, such as in an ongoing conversation or shared endeavour, refers to *contingency*.

The simplest definition for *spontaneous communication behaviour*, as used in this project, would be a behaviour that has not been prompted. It is "communication in the absence of [a] defined antecedent" (Chiang and Carter, 2008, p.695) or "communication in the absence of a specific instruction to perform the behaviour" (Carter, 2002, p. 169; in reference to Sigafoos and Reichle, 1993). Spontaneity forms the core of how *initiation* is defined in this project: a behaviour that is unprompted, socially directed, intentional, and begins a new interaction. Indeed, Chiang and Carter (2008) treat the terms *spontaneous* and *initiation* as effectively interchangeable. Initiations are "new" interactions in that they are not contingent on the social partner's previous action, that is, they have no defined antecedent in the immediate past communications between those particular partners. Prompts, contingency, and antecedents all refer to the same underlying concept. A prompted child behaviour is contingent on that partner's action. Thus, *responses*, as the opposite of initiations, are behaviours that require antecedents. They are contingent on the social partner's previous behaviours, and are cued by the form and content of those behaviours.

Identifying initiations in extended, reciprocal interactions is trickier. While it would be perfectly reasonable to consider the "opening move" to be an initiation and all subsequent moves to be responses (because of their direct antecedents), this is not very informative. It would not be clear who was driving the interaction to continue, for example by asking a follow-up question or introducing completely new information. The current view of initiation follows Breen and Haring, who say "An initiation [is] defined as any behavior, verbal or nonverbal, that began an interaction. A new initiation [is] coded if there [is] a change in topic or a mutual break in focus followed by a refocusing of attention" (1991, p.339). Taking this view aims to give a child maximum credit for his or her role in continuing and developing the interaction.

2.3.2 Initiation behaviour in autism

While communication difficulties are rightly identified as central to the concept of autism, not all communication appears to be equally difficult for people on the spectrum. Responding to others' verbal and non-verbal bids for interaction appears to—generally—be less impaired compared to TD children than is the initiation of new bids for interaction. For this project, the main focus is on initiation of communication, so difficulties with responding will not be reviewed in depth. Indeed, when these are discussed in the literature they are often treated more generally, as a part of overall communication difficulties.

Children on the autism spectrum tend to initiate communication infrequently (compared with TD peers or even developmentally delayed peers) and their initiations are more likely to be imperative than socially motivated[6]. The coarsest distinction between initiation functions is between initiations that are *imperative* and those that are *non-imperative*. White and colleagues (2011) give a pleasingly simple definition of *imperatives* as "acts that function to gain access to an object" (p. 1284) and, more abstractly, as "tangibly maintained behaviours" (p. 1294). This second definition makes clearer that imperatives are not only concerned with physical objects: another person performing a desired action like hugging or starting the DVD player (or leaving the child alone!) can also be tangible maintenance. *Non-imperative initiations* or *social initiations* are those that are more socially motivated: they are not trying to obtain a physical thing or directly bring about a desired state of affairs. Their maintenance, such as it is, is the communication or interaction itself. Examples would include greeting, sharing affect, or commenting. The term "non-imperative" is used interchangeably with "social" throughout this document (in relation to types of initiations). Some initi-

[6]Here, "children with autism" refers to those who are closer to the target group in this project (see participants sections of Chapters 5 and 8), who are not considered to have high-functioning autism or Asperger syndrome. Children in those groups may show a very different communication profile, and indeed may struggle to *limit* their initiations to appropriate times, places, and topics.

ations may be neither strictly imperative nor social, such as those that attract attention to a person prior to speaking to them or making some other communication.

Observation work in naturalistic settings has concluded that imperative communications considerably outnumber non-imperatives for children in this group (e.g. Chiang, 2009; Stone and Caro-Martinez, 1990), though these papers are not particularly clear about the types of events and communication opportunities present while the children were being observed. Other experimental work in this area suggests that imperative initiations are (relatively) intact in autism, compared to initiations that involve attention sharing (Goodhart and Baron-Cohen, 1993; Loveland and Landry, 1986). This imbalance appears to persist at least through late childhood (Loveland and Landry, 1986). This said, it is important to keep in mind that the majority of children with autism *can* and *do* initiate to people around them. Children with autism may initiate readily to manage their own environments and others' behaviour (imperative functions), for example by requesting or protesting objects or activities, and using adults as "tools" for tasks they cannot do themselves. SCERTS[7] researcher Emily Rubin humorously but accurately describes this last function as being a "human can opener"[8]. As Wetherby remarks, "it is not that children with ASD do not communicate but rather that they do not readily communicate for social goals or purposes" (Wetherby, 2006, p. 9).

This binary imperative/non-imperative distinction is the level at which initiations are generally discussed in this proposal, where their functions are discussed at all. The current work is primarily concerned with "initiation" as a general category of communication behaviour; systematic analysis of initiation functions is reserved to future work (Chapter 9). However, the imperative-social distinction is introduced here to help fill out the picture of communication difficulties in autism, and clarify that not *all* initiation is equally difficult for people with ASC.

2.3.3 Initiation and links to long-term outcomes

From one point of view, the value of initiation is simple: an individual who can begin new interactions with a range of social partners can *do* much more than one who only responds to others. With responsive skills, a person can make a choice—if it is offered—or participate in a conversation or activity—if invited. Someone who is able to initiate in a relevant, functional way can better express needs and preferences, protest

[7]The SCERTS model, Prizant et al. (2005)

[8]Verbal presentation at SCERTS model training session, Birmingham, UK, 21st September, 2010.

or request things, ask to join an activity, share something, volunteer information, and more. S/he has greater agency. Any degree of initiation skill, verbal or nonverbal, allows a person to have more control over him or herself than s/he would have if only responding to others. Acquiring *and using* a greater range of skills increases these benefits. Emphasising *use* of initiation skills is important. Especially in autism, there may be a divide between what skills children are *able* to use under particular circumstances (for example in an assessment or therapy setting), and what skills they regularly and independently use in daily life. The assessment part of the SCERTS model (Prizant et al., 2005, Volume 1, Assessment) is one example of an intervention programme that tries to capture this difference, scoring behaviours based on the degree of regularity and independence with which they are demonstrated, and across how many settings. They acknowledge a middle level of acquisition, in which children can perform a behaviour, but do so occasionally or inconsistently, only with support, or in a single type of activity[9].

Discussion of communication initiation and autism, particularly in relation to other skills, frequently focuses on initiation of *joint attention*. Joint attention is often explained as a triadic social coordination between two persons and an object or event in the environment (e.g. Bakeman and Adamson, 1984; Charman, 2003; Walden and Hurley, 2006; Whalen and Schreibman, 2003; Mundy and Newell, 2007; Vismara and Lyons, 2007), in which it is necessary to monitor another person's attention in relation to one's own. Perhaps one of the best descriptions of triadic coordination comes from Donald Davidson[10]:

> "The basic situation is one that involves two or more creatures simultaneously in interaction with each other and with some aspect of the world they share... It is the result of a threefold interaction, an interaction which is twofold from the point of view of each of the two agents: each is interacting simultaneously with the world and the other agent" (2001, p.128).

A key word here is simultaneous, which highlights why many definitions of specific joint-attention-constituting behaviours (e.g. in the ESCS manual; Mundy et al., 2003) require the co-occurrence of an object-directed and a partner-directed action, such as gaze to the partner's face while touching a toy. Merely attending to another person's current focus of attention is insufficient—the jointness is missing. Neurotypical humans frequently orient to interesting objects or events in the environment because oth-

[9] See (Prizant et al., 2005, p.135) and the assessment forms appendix in the same volume.

[10] He actually uses the term triangulation to refer to this concept, rather than triadic coordination, writing in the context of the emergence of propositional thought.

ers are already looking at them, and do this even in the absence of deliberate direction. However, in order for attention to be coordinated, the person directing attention must do so *intentionally* (Tomasello, 1995). The initiator must be aware of the respondent and the respondent's focus of attention in addition to the object or event to which he himself attends. Conversely, the element of coordination is missing if both social partners are attending to the same object or event but not to each other. Thus, *initiating joint attention* means beginning one of these episodes of coordinated attention but deliberately monitoring a partner's attention, and directing it to something in the environment.

Joint attention is not an isolated skill, but rather a bundle of behaviours that are built on other developmental experiences such as orienting to social stimuli (e.g. Dawson et al., 2004; Butterworth and Jarrett, 1991) and in turn enable the development of further key skills. Charman (2003) gives a good overview of the evidence for joint attention as a "pivotal skill". Longitudinal studies of children with ASC suggest that early joint attention abilities (both in initiation and responding) can be highly predictive of language outcomes later in childhood. The "coordination" part of joint attention is an early version of maintaining contingency throughout a conversation or cooperative task. Some researchers suggest that it may be part of the skills that enables children to acquire language, "link[ing] spoken words with the objects and events to which they refer" (Walden and Hurley, 2006, p. 280). Better performance in joint attention during the preschool years appears to boost both expressive and receptive language development over time. Mundy et al. (1990) found that non-verbal, declarative joint attention such as pointing and showing objects[11] significantly correlated with language ability at an initial assessment and at a one-year follow up. More importantly, out of the various communicative behaviours assessed, only joint attention scores *predicted* language development over that period. A more recent study from Bono et al. (2004) found a modest correlation between frequency of joint attention initiations and gains in preschool-aged children's language over a one-year period. In a rare randomised control trial and follow-up study in this area, Kasari and colleagues (2006) delivered six-week joint attention and play interventions[12] to a group of children with autism aged 3-4 years, and found that both interventions produced significant skills gains. Those in the joint attention group showed significantly more joint attention initiation

[11]Based on a combination of measures in the early version of the Early Social Communication Scales (ESCS; Seibert et al., 1982)

[12]NB: these were two separate interventions, not an intervention with both joint attention *and* play skills.

and responding, in a post-intervention structured task and interaction with a parent. The five-year follow-up study (Kasari et al., 2012) found that children's early joint attention initiation ability (in the 2006 study) was one of the key predictors of later functional verbal language, for those children who acquired it[13], along with higher levels of play, and participation in the one of the experimental (intervention) groups versus the control group.

In a study from Charman and colleagues (2003), children's scores on joint attention at age 20 months, as operationalised by gaze shifting between an adult and a toy that had ceased working, were positively linked to their receptive language development at age 42 months. There did not appear to be a relationship between this particular measure of joint attention and expressive language. However, a study by Sigman and Ruskin (1999) showed the opposite pattern, with joint attention ability related to expressive, but not receptive, language gains. Other studies clearly suggest that it is actually joint attention *response* that is most important for language development, and most predictive of that development over time (e.g. Sigman and McGovern, 2005). For further examples, Yoder and McDuffie (2006) give a good general discussion of the evidence related to the proposed relationship between the various types of joint attention and language ability.

Recent work suggests that not all joint attention initiations (IJA) may be equal when it comes to their contribution to later skills. Pickard and Ingersoll (2015) point out that IJA have often been divided into high-level and low-level behaviours in assessment (e.g. Mundy et al., 2003), but may be lumped together in studies of skills and outcomes—perhaps leading to the rather mixed results. High-level IJA include pointing, showing, and other actions that indicate *sharing* of affect and experience (i.e. a social motivation; Mundy and Crowson, 1997). Low-level IJA include gaze-shifting and similar behaviours, that do require attentional monitoring but may not involve the same level of *communication*. Per their own results, Pickard and Ingersoll (2015) suggest that only high-level IJA behaviours are associated with language and imitation, and that these behaviours may be the more important precursors of advanced social communication.

This assortment of results suggests that, even if the exact relationships between types of joint attention and types of language outcomes are not fully untangled, there *is* overall a positive relationship between joint attention and language. Gaining functional language, especially by early school age, is associated with improved social, adaptive,

[13] Approximately 80% of the follow-up sample.

and intellectual outcomes into adulthood (e.g. Venter et al., 1992). Moreover, targeting joint attention may have collateral benefits in other areas of functioning, and in children's quality of life. For example, Whalen et al. (2006b) attempted to teach joint attention response and protodeclarative pointing, but they also found positive changes to do with imitation, play skills, and spontaneous speech in general. Kasari and colleagues (2010) found that a parent-mediated joint attention intervention, as part of a randomised control trial, not only improved children's responsiveness to joint attention, but also produced gains in joint engagement with the parent, and diversity of their functional play. These gains were maintained at a one-year follow-up. Findings of this type further support claims that joint attention and IJA in particular are pivotal or foundational skills for autism, affecting many other areas.

In closing, some of the most convincing evidence for the importance of communication initiation are the researcher-developed policy documents and intervention strategies that, having reviewed large bodies of evidence, treat initiation as central. Spontaneous communication has been identified as a National Research Council (USA; NRC, 2001) instructional priority for autism, in their landmark report reviewing 20 years of literature on autism intervention and services. This effectively means *initiation* of communications, not dependent on a partner's prompting or cueing. Indeed, in their executive summary, the NRC lists "functional spontaneous communication" as a key recommended priority within educational services (NRC, 2001, p.9). Social initiations are again highlighted in its expanded discussion of educational and intervention objectives for children with ASC; indeed, the authors go so far as to say that "functional, spontaneous communication should be the primary focus of early education [for autism]" (p.221). The SCERTS model (Prizant et al., 2005), following the NRC report, treats "functional, spontaneous communication" as a main priority; this is one of its three domains of intervention. Other best-practice or strategy documents repeatedly refer to social communication more generally, without specifying initiations and responses as targets (e.g. Parsons et al., 2009b; The Scottish Government, 2013).

2.3.4 Approaches to social communication intervention

Given their centrality to how autism is defined and diagnosed, many behavioural intervention programmes have focused on acquisition and use of social and communicative skills. *Behavioural interventions* are deliberate alterations to a person's environment or experiences with the goal of changing[14] their behaviour, leading to some developmental outcome such as mastering a daily-life skill or improved use of verbal language (Walden and Hurley, 2006). This section gives a very brief survey of some approaches to communication interventions for autism. It is meant primarily to highlight the different views and underlying assumptions about communication and learning that are present in this area, rather than to present evidence for their effectiveness. These more traditional, "face-to-face" approaches give context for the intervention strategies that seek to promote communication through environmental change, and many of the technology-based interventions discussed in a later section.

Intervention philosophies vary hugely. The major approaches differ in the naturalism of their interactions, the style and frequency of adult prompting, reinforcement methods, emphasis on behavioural forms versus functions, and the degree to which they are child- or adult-led. At one extreme are strategies like *discrete trial training* (DTT; e.g. Lovaas, 1987). These programmes are highly-structured and adult-directed, and often depend on arbitrary reinforcement for behaviour. DTT teaches small "pieces" of behaviours or language through massed trials, often consisting of repeated adult prompts and child responses. As Yoder and McDuffie (2006) point out, the core assumption of this approach is that learning the form of a behaviour, will lead to the child learning its function (as demonstrated by the adult). Approaches of this type can create measurable behaviour changes, but there is always the risk that their arbitrariness may lead the child to develop a *therapy register*, or repertoire of behaviours produced during a teaching session, but not produced otherwise. Frequently, DTT is a "full-time" intervention, with a child working one-on-one with a therapist up to 40 hours per week.

At the other extreme from DTT are child-led approaches such as *responsive teaching* (RT; Mahoney and MacDonald, 2003) and Floor Time (e.g. Wieder and Greenspan, 2003), which emphasise entering in to the child's world as much as possible and following his or her lead in terms of which activities are pursued, and in what manner. In this view, the adult social partner's availability, responsiveness, and positive rapport with the child are all-important. Building shared engagement is a foundation for more complex activity. Unfortunately, there is not extensive research on how such approaches are implemented and assessed. Harjusola-Webb and Robbins (2011) give one recent example of a multiple-baseline study that does report an increase in various

[14]"Change" is meant broadly. This may mean acquiring a totally new behaviour, extending a behaviour to a new context, or substituting a new behaviour for one already known.

communicative behaviours when RT is implemented in the classroom, but it does not say whether the increases are significant.

Intensive interaction can also be considered a child-led approach, and one that may be particularly relevant for individuals with significant learning disabilities in addition to ASC. Initially developed by Nind and Hewett (e.g. Nind and Hewett, 1994), it targets the very foundations of social interaction, such as physical proximity and contact, and engaging in shared attention and activity. Intensive interaction draws on concepts from the infant-caregiver interaction literature, and emphasises the skilled communication partner (i.e. the adult) responding and adapting to the child, imputing intentionality to him or her, and following his/her attention and actions with the goal of creating mutual pleasure and interactivity. Case studies (e.g. Kellett, 2000) and small-n studies (Nind, 1996) provide some evidence for Intensive Interaction having increased early social and communicative behaviour in children with autism and severe learning disabilities (e.g. contingent vocalisation, use of gaze). The approach has become sufficiently widespread in the UK and is considered sufficiently-well evidenced to be highlighted in an official government policy document on learning disability, in relation to "facilitating meaningful two-way communication" (Department of Health, 2009, p.37). However, as Hutchinson and Bodicoat (2015) point out in their systematic review, the overall picture of evidence is unclear and difficult to evaluate, due to small studies, population heterogeneity, unclear reporting, and problems controlling for sources of bias.

In the middle between these two poles are *transactional approaches*, such as the SCERTS framework (Prizant et al., 2005, 2003). These approaches use natural, daily-life situations, and see child-adult interactions as a key mechanism of communication development. In a major contrast to DTT-like strategies, transactional approaches are "based on the belief that children acquire language and communication by using it, not through practising its separate parts" (Quill, 1995a, p.165). Here, the adult social partner responds to the child's communicative actions, following their attention or play in order to scaffold turn-taking, sharing, and other social activities. They teach through game-like strategies with graduated levels of prompting, with the goal of eventually making functional communications and social interaction rewarding for their own sake. Unlike in DTT, the adult uses prompts only *after* the child has initiated communication and shown a strong motivation to continue their engagement. In all cases, a prompt should be at the minimum and least invasive level required to encourage the desired social behaviour, and should be faded out as the child comes to consistently respond to that level or type of prompt. These preconditions for prompting are meant to decrease the likelihood that the child will become dependent on prompts, and to increase the chances that the child will come to associate his communicative efforts with positive affect and self-efficacy.

As a final note, it is important to keep in mind that many intervention programmes and research projects (including ECHOES) do not strictly adhere to any one approach, but mix them together as needed, based on the goal and the medium of content delivery. A service setting, such as a school, or an individual practitioner may also mix approaches to meet child needs or operate within the constraints of the setting. Thus far, technology-based communication intervention tend to offer more embedded teaching and plenty of scaffolding, rather than repeated practice (though see Whalen et al., 2006a, 2010, for exceptions).

2.4 Extrinsically motivating initiation

2.4.1 Motivation and the environment

Social communication, particularly joint attention, may not be intrinsically motivating to a child with ASC in the same manner (or to the same degree) that it is for a TD child (e.g. as discussed in Wetherby, 2006; Leekam and Ramsden, 2006). It is possible that some children with ASC are motivated by the tangible rewards of communications that allow them to gain access to objects or assistance from others (i.e. imperative communication), but not by the social rewards of sharing-type behaviours. This suggests that if the social component of information sharing is not (sufficiently) *intrinsically motivating* for a child to initiate those behaviours spontaneously, the motivation to share or comment needs to be *extrinsic*, out in the world.

There is a long tradition of eliciting target communication behaviours (particularly verbal ones) using arbitrary but attractive extrinsic motivators (e.g. snack food, tokens). *Applied behaviour analysis* (ABA) uses such strategies heavily (e.g. Charlop et al., 1983, 1992; Karsten and Carr, 2009), as does DTT (Lovaas, 1987), even in recent work (e.g. Fisher et al., 2014). These types of motivators are arbitrary in the sense that they are unrelated to the behavioural form that is being taught and also do nothing to teach the child its function. While such methods may motivate the child to produce the desired behaviours, there is little chance that the behaviours will ever become spontaneous and self-sustaining outside of the teaching situation or will generalise to other contexts, due to the dearth of conveniently positioned, token-dispensing adults in the outside world. Moreover, programmes of this type tended to be clinician-driven in terms of what is chosen as a priority for teaching, rather than following the child's interests.

Having ruled out *arbitrary* extrinsic motivators as unlikely to produce self-sustaining communication behaviours, that leaves motivators that are *inherently related* to the type of communicative behaviours that we would like to encourage the child to practice. Unfortunately, the highly structured, timetabled, predictable environments that may support children with ASC and seek to reduce their anxiety may also limit the extrinsic motivations for them to communicate. Needs may be met by carers "reading" the child, rather than the child needing to actively communicate. However, if an environment can effectively remove the need for communication, it could also be altered in the opposite direction, to become "worth communicating about" in a way that it was not before. One such extrinsic motivation strategy has been to increase or vary the communicative demands that are already present in the child's environment and activities, further discussed in the following section.

2.4.2 Altering environments to motivate communication

The idea of introducing surprises or disruptions into a child's environment in order to motivate communication is not new, although it does not appear particularly often in the *research* literature—it is largely addressed in books for clinicians, or parents and carers. This general concept is in no way specific to autism. It is generally presented for the purpose of enhancing communication quantity/quality or creating a greater necessity for communication in daily activities and routines. McClenny et al. (1992) give perhaps the most explicit statement about the assumptions underlying this general approach, noting that when daily routines and well-established schemata are altered, "these modifications may force the child to realise something is amiss, and *create the need for the child to protest or make a request*" (p. 231, emphasis added).

The earliest sources are more generally concerned with speech and language therapy, rather than autism. Later sources frequently repeat and slightly extend the ideas of earlier sources, but without substantially developing or evaluating them. There does not seem to be a consistent vocabulary for this idea, and from inconsistent or absent citation between authors, it is possible that this idea emerged from clinical or practical experience several different times. Overall, the information on this topic is most frequently presented as suggestions for practice. There are limited anecdotes of actual children reacting to events of this type, and extremely limited empirical evaluation. The latter employ this type of strategy as part of a larger programme. Given the limited literature and the inaccessibility of some early papers, it is worth briefly tracing how this idea has been presented, and for what purposes. To avoid repetition, a composite list of suggestions for environmental changes appears after the chronology.

2.4.2.1 Prior work on unexpected events to elicit or increase communication

The earliest appearance of this idea—as far as could be determined—seems to be in Constable (1983), who is interested in "nonlinguistic events that increase the likelihood that the child will initiate a specific speech act" (p.109). She suggests violating routine events, hiding or withholding objects, or violating their function. McClenny et al. (1992) pick up the ideas and goals of Constable (1983), but empirically evaluate their effectiveness with developmentally delayed young children (discussed separately below). Chronologically after Constable (1983), Snyder-McLean et al. (1984) refer to a similar idea, in a discussion of varying *joint action routines*, or ritualised, reciprocal interaction patterns, such as a play sequence with defined roles. To support generalisation, they suggest "deliberately introduc[ing] an element of surprise to evoke a spontaneous declarative or interrogative communication act" and suggest forgetting some critical object for a routine as an example of this (p. 218). This same year, Wetherby and Prutting (1984) mention several similar situations as a part of an assessment of the communicative profile of children with ASC, without the researcher verbally prompting the child. These strategies are re-branded as *communicative temptations* and substantially expanded in Wetherby and Prizant (1989). The authors describe these as "creating opportunities that entice specific attempts at communication" though they note that each temptation may, in fact, elicit a range of communicative functions. Here, their goal is to assess children's communicative intent in a clinical context, rather than to promote communication in a natural context. Interestingly, neither Wetherby and Prutting (1984) nor Wetherby and Prizant (1989) cite either of the earlier two sources, despite the similarity of the ideas.

Several more practice- or parent-oriented sources pick up these ideas in the 1990s[15], now focusing clearly on increasing or improving daily-life communication, not on assessment. Howlin (1998) gives a number of anecdotal, clinical-practice examples of

[15] In addition to the experiments in McClenny et al. (1992).

how "routines, or rather the disruption of routines, can be a very effective means of increasing communication" for children with autism (p. 122)[16]. Twatchman (1995) takes a slightly different view of this same general idea, presenting environmental changes as "staged problem-solving situations" for children with autism (p. 152). Both views tap into the idea of altering the environment so as to embed demands for communication in contexts, and stress that these contexts should be *already motivating* to the child in some way (not a concern of earlier authors). They also tie in closely to the more general body of work on *naturalistic language teaching*[17], which embeds teaching interactions into the child's customary activities and uses stimuli that are already present in the space. Naturalistic strategies are child-led, in that the adult only offers a situation-appropriate prompt after the child has already initiated an action, such as reaching for a toy[18].

More recently, several general programmes for autism support or education have drawn on these earlier sources for specific intervention strategies. The Hanen 'More Than Words' (HMTW) parent-implemented intervention programme manual (Sussman, 1999) has an entire section dedicated to "Giv[ing] your child a reason to communicate" (p.68-83), which acknowledges Wetherby and Prizant (1989) as the main source of its suggestions. HMTW is targeted at parents, and (unlike the original source) focuses on daily-life routines and playtime. Of the various how-to resources, this is the one that most clearly illustrates possible strategies, and mixes specific examples (e.g. pretending not to hear the doorbell ring, giving the child an opportunity to point it out, p.81) with more general concepts that could help parents generate their own ideas, such as using "people toys" that are attractive but need adult help to make them work. HMTW has not been extensively evaluated, but a randomised control trial (Carter et al., 2011) comparing the programme to a "business as usual" control group did not find significant effects for child communication across the group, though they did note positive effects for some participants.

The SCERTS programme (Prizant et al., 2005), on which Wetherby is the second author, incorporates her earlier work on "communicative temptations" as a tool in its assessment procedures, as a way to try to elicit certain types of intentional communication[19]. This is part of gaining a sample of the child's behaviour, or potential behaviour,

[16]Howlin also includes similar content in her later writing (2006).

[17]"More general" in that these strategies have never been specific to autism intervention.

[18]See Rogers (2006) for an overview of the history and characteristics of naturalistic language teaching approaches, specifically as applied to the autism spectrum.

[19]See pages 157-159, SCERTS volume 1.

and are advised as a supplement to naturalistic observation. Some of their suggested "temptations" are more openings for communication, such as waiting for the child's turn in a predictable turn-taking interaction. Some others are more like surprising, unexpected, or disruptive events, for example to "suddenly spill some substance or do something unexpected in front of the child" or to "end a preferred activity abruptly or unexpectedly" (p.159, Table 7.5)[20].

In a school context, Kossyvaki et al. (2012) have targeted adult intervention styles as a means to increase the spontaneous communication of children with ASC, using both general principles and specific "communicative opportunities". A number of these opportunities are in line with the earlier work on environmental changes and routine disruption[21] such as "give material the child will need help with" (e.g. a box lid that s/he cannot remove alone; p. 177). The school staff members did increase their use of these strategies during the intervention (per analysis of videoed interactions), and children also increased their use of initiations. However, as the intervention included a wide range of adult interaction strategies introduced as a package, the evidence for the contribution of the "communicative opportunities" (versus other strategies) is at best indirect.

Suggestions for environment-altering events to motivate child communication

The following list of suggestions is drawn from across the sources mentioned above. It does not give every example, but is meant to represent the range of ideas present.

- Giving a child inappropriate or mismatched objects: Using a spoon too large for a food container or batteries that are too large for a flashlight (Layton and Watson, 1995).

- Switching things: a familiar wind-up toy for a broken version of the same toy, or labels on containers for snacks so that this is not apparent until after they are opened (Layton and Watson, 1995).

- Creating visible but inaccessible objects: place a desired item out-of reach (Constable, 1983; McClenny et al., 1992; Kossyvaki et al., 2012), or present a con-

[20]Prizant et al. (2005) also make a useful and unique contribution in that they *also* give suggestions for situations to elicit children with more advanced (i.e. conversational) language (see page 160). None of their listed suggestions represent routine-disruption or similar, possibly because this type of strategy may be less appropriate for a more verbal, more able, or older child.

[21]The authors cite HMTW (Sussman, 1999), SCERTS (Prizant et al., 2005), and personal experience in practice settings.

tainer with desirable contents that the child needs help to open (Layton and Watson, 1995; Wetherby and Prutting, 1984).

- Violating object function: trying to use the wrong object for a task, or to use a broken object (Constable, 1983; McClenny et al., 1992).
- Making substitutions: Substitute a task's customary object for another suitable object (Prizant et al., 2005, p. 158).
- Use of attractive "people toys": these require adult help to operate, such as wind-ups, blowing bubbles, spinning tops, or balloons (Sussman, 1999).
- Withholding objects or attention: giving part of the items needed for a task, stopping during an interaction (Constable, 1983; Layton and Watson, 1995; Kossyvaki et al., 2012). Wetherby and Prutting (1984) specifically suggest "eating a consumable item that the child likes in front of the child without offering the food to the child" (p.367).
- Adults making deliberate mistakes: dropping, forgetting or apparently being unaware of things. Sussman (1999) calls this "creative stupidity" (p.79).
- End (desirable) activities suddenly or early (Prizant et al., 2005).
- Take advantage of when things naturally go wrong! (Sussman, 1999).

It is important to consider the type of communication that these approaches hope to elicit or increase. Given these situations, what does the child actually communicate about? What goal is being achieved? It is fairly obvious that most of these situations are most likely to encourage imperative communication, to get objects or control others' behaviour (e.g. eliciting help). The anecdotal, practice-based examples of routine-disrupting or unexpected events (versus those mentioned in the context of assessment) frequently leverage well-known motivators for all young children: gaining access to snacks and attractive toys! While likely to be effective, these are not ideal conditions for supporting socially-motivated initiations, such as showing objects, sharing affect, or commenting. Some of the single-child anecdotes include instances of non-imperative communication, such as the boy who comments "no peanut butter" after finding out that the usual jar is full of frosting (Layton and Watson, 1995), but there is no good information about how prevalent this might be. Such instances are encouraging but are apparently very much in the minority, which is not surprising when many authors have explicitly targeted protesting and requesting.

2.4.2.2 Experimental investigation of unexpected events and communication

McClenny et al. (1992) are a rare example (perhaps the only example so far) of investigating unexpected events as a stand-alone variable in an experimental setting rather than as part of a particular treatment programme. In this case, their focus is on child communicative responses to unexpected events, from developmentally delayed preschool children[22]. The experimental set-up includes one child and one experimenter engaged in a session of two water-play activities (one activity with unexpected events and one as a control). These were open-ended and did not resemble a daily-life task (e.g. washing something). The experimenter's role and utterances were closely scripted, and included making false assertions (e.g. making statements about the child's current play but using the wrong verb or object name), temporarily withholding objects from the child or hiding them from view (plus a relevant statement), or atypical object actions[23] or using atypical sequencing of steps in a routine. The control condition used the equivalent correct (expected) information statements, sequences, and objects.

Responses in the unexpected event condition were more than twice as frequent than in the control condition, with the greatest response frequency to "withhold objects" events, and then to violation of object function. Responses were subdivided into different "functional categories" of comment, request, or answer/acknowledgement[24]. What is perhaps more important and more interesting than the increase in response in the unexpected condition is that, unlike the other examples of environment modifications, children's communication cannot have been motivated by gaining access to necessary or preferred items. Even in the case of "withheld object" events, the child received the toy after a short pause regardless of whether s/he made a response.

In several ways, this paper comes nearest to being a direct precursor to the ECHOES observations and analysis reported here. Unfortunately, researchers do not appear to have directly built on this research and few later sources cite it.

[22]Without ASC, but close to ECHOES participants in terms of developmental age and language ability.

[23]As an example, the experimenter might try to write with the toy shovel and say "I'm writing with my shovel" (p. 244).

[24]The definitions were extremely broad, and appear to mix imperative behaviours such as "solicits an action [or] object...from the listener" (p. 245) with those that are likely to be socially motivated, such as "direct[ing] the listener's attention to some observable referent" (p. 245). The authors also focus on verbal language, and do not report any non-verbal behaviours.

2.5 Technology and autism

For autism researchers and practitioners, or for technology researchers with no expertise in autism, encountering ASC-tech projects raises an obvious question: why apply technologies to autism? What is the rationale for bringing these things together? The very short answer is that interactive technologies are perceived to have useful affordances that can be a good fit for the needs of people on the autism spectrum, and in some cases to offer functionality or support that face-to-face techniques may not. The following section aims to review some of this reasoning, and examine several specific ASC-tech examples that have sought to teach or support social communication skills. As much as possible, it connects these examples to the more general approaches to communication intervention, discussed earlier in the chapter. Chapter 3 returns to ASC-tech at more length, with a higher-level discussion of its state as an emerging, interdisciplinary field, the range of issues it has tried to address, and some of its current challenges.

2.5.1 Benefits and affordances of technologies

While interactivity and experiential learning are one of the strongest arguments for using a wide range of technologies to teach or practice skills, the research literature argues that technologies have a number of additional benefits specifically for users on the autism spectrum, mainly to do with the fact that they are controllable, repeatable, and thus predictable. These potential benefits are summarised below. While often originally presented in discussions of virtual environments for autism, the points seem equally applicable to related technologies, such as 2-D games or mobile apps. Most of the following points are repeated in multiple papers, but specific citations are given where relevant[25].

1. *Technologies can offer a controlled, personalisable environment*

    - Screen-based technologies, such as a virtual environment presented on a desktop computer, help to avoid sensory and attentional overload by "limiting" the child's attentional window (Schmidt and Schmidt, 2008).
    - Many distractions can be removed in terms of unnecessary "background"

[25] A concise general discussion of the intervention potentials of virtual reality may be located in Schmidt and Schmidt (2008).

sounds, objects, etc. It is also possible to remove objects that are of particular interest to the child and may prevent focus on the target activity.

- Sensory and processing needs and preferences can be accommodated (e.g. colours, sounds, overall pacing, level of language).
- Screen-based technologies can capitalise on the visual processing strengths of people with ASC (e.g. Quill, 1995b, and as briefly reviewed in Shane and Albert, 2008). Methods such as video modelling particularly capitalise on this perceived strength.

2. *High predictability and consistency are possible*

   - Overall, there is capacity for extensive repetition of activities and interactions, with great consistency between them.
   - Virtual environments allow non-threatening, repeatable practice of situations that may cause the child anxiety or hold real-world danger, particularly situations with a social component.
   - Virtual characters are less socially demanding and more predictable than human social partners, and also capable of endless repetition (without impatience!).
   - Overall, technologies may better meet the "need for sameness" in ASC than do face-to-face supports or teaching.

3. *Content and delivery are adjustable*

   - Potential for automated scaffolding for each user, with program difficulty increasing or prompts being withdrawn as the user improves (i.e. along the lines of many strategies in the intelligent tutoring systems literature).
   - The content or setting can be modified (without building a new environment) in order to encourage skill generalisation. See Parsons et al. (2004) and Parsons et al. (2006) for examples.
   - There are many options for input devices to suit users' age and motor skills.
   - Automated content can be combined with a Wizard of Oz interface for formative testing of a later-automated system, or for an intervention (e.g. Tartaro and Cassell, 2008).

4. *Motivation*

   - Technologies are widely identified as being motivating for many people with autism (partly because of the predictability and repeatability identified above) and may lead to willingness to continue participating in an activity or intervention over time.
   - Surveys of technology use on the autism spectrum (e.g. Fletcher-Watson and Durkin, 2015; Shane and Albert, 2008; Putnam and Chong, 2008) identify technology use as a popular leisure activity, across a wide range of devices and particular tasks or programs (games, social media, watching/uploading videos, etc.).
   - There is the potential to incorporate an individual's special interests, a strategy used successfully in a face-to-face intervention by Vismara and Lyons (2007).

5. *Practical affordances*

   - Technologies may be able to augment, extend, or reinforce existing face-to-face intervention (e.g. extra skills practice at home).
   - Technologies may be able to extend or provide new provision in areas where professional face-to-face support for autism is effectively unavailable due to lack of specialists, intervention costs, or geography.
   - People with ASC can carry mobile technologies between contexts, to access their support on-demand.

Technologies may indeed be particularly suited to users with autism for the reasons mentioned above, but it is possible that the current line of work may eventually add different reasons to the list: their potential to delight, amuse, surprise or challenge the user in ways that may be difficult to achieve in his or her daily, physical environment. The ECHOES project and other technologies targeted at users on the autism spectrum have barely scratched the surface of the events and effects that are possible in a digital medium. Given that the surprises in ECHOES were largely accidental and initially viewed as a liability by the neurotypical adult researchers, there seems to be considerable scope for deliberately extending these aspects of interactive technologies and shaping them into a tool for future interventions.

2.5.2 ASC-tech for communication

To illustrate some of the work done in ASC-tech on promoting communication, the rest of this section describes two examples of work that share key features with the ECHOES project and new *Andy's Garden* games. The examples discussed here do draw on general intervention approaches, but are not direct implementations of face-to-face or paper tools. The descriptions aim to give a sense of what each technology was like, and how it represents an approach to using ASC tech to support communication. A few key findings are reported, but the authors' goals and approach are of more interest. The two examples are followed by additional commentary. For a broader review of work on ASC-tech for communication, see Wainer and Ingersoll (2011). A number of studies in the ASC-tech meta-analysis from Grynszpan et al. (2014), and the reviews from Fletcher-Watson (2013) and Goldsmith and Leblanc (2004) also address social communication skills.

Virtual peers to scaffold contingent discourse Tartaro and Cassell (2008) are interested in using *virtual peers*, or "life-sized computer-animated, interactive children" as a means to scaffold the contingent dialogues of children with ASC, through collaborative storytelling (p.382). Their overall approach to this issue is broadly transactional or naturalistic, using interaction between a child with ASC and a more experienced or able partner as a mechanism for development. A friendly, positive atmosphere between the target children and their peers also seems to be quite important. While their setting is experimental, it bears a clear resemblance to the type of situation children may encounter in school or at home, with attractive, off-the-shelf toys and an available peer.

In their study, children with autism played and told stories with Sam[26], a virtual peer who appeared close to their own age range (7-11 years). Sam is projected at life size onto a large screen, behind a wooden doll-house with related figures and objects. Part of the house is also projected with Sam, making it appear to extend into her space as well. To study the effect of the facilitation strategies, the authors compared play with Sam to play with a TD peer. In both conditions (counterbalanced), children were given a deliberately minimal instruction, inviting them to make up stories using the toys. In the Sam condition, they were additionally introduced to her and told she "loves" to tell stories. A researcher controlled Sam with a Wizard of Oz interface, directing Sam

[26]The authors explain that Sam is deliberately of ambiguous gender; this is likely meant to make him/her relatable to the largest number of children. In the paper, they refer to Sam as *she*.

to use pre-recorded speech and animated gestures. It is important to emphasise that neither the virtual peer set up (technology component), the TD peer, or the toy set-up are trying to explicitly *teach* anything, or complete a *specific* activity.

Video data of the interactions was transcribed and annotated for contingency and topic management. Children with ASC showed less contingency overall with both types of partner. However, their contingent utterances and use of topic management both *increased* over the course of a 15-minute interaction session with the virtual peer—an effect not present during the sessions with TD children. Overall, the findings in Tartaro and Cassell (2008) suggest that the virtual peer's pattern of "set[ting] up narrative situations" makes a difference, increasingly scaffolding the communication of children with ASC over a short session (p. 387). The specific interaction strategies assigned to Sam (and deployed via the Wizard of Oz interface) appear to have been effective in supporting contingent communication[27]. As the authors point out, this change in contingency *over* the session may be related to children needing to familiarise themselves with the situation and with Sam; it raises the possibility that longer or multiple sessions may be able to do even more. In full, the study results are convincing proof-of-concept for a future intervention.

This study is also an excellent demonstration of a technology that acts as an *indirect* and open-ended mediator of interaction between a child and an adult. Sam is not automated, but controlled by a researcher in a way that is likely not apparent to the child. A successful interaction takes place *through* her. The researcher uses his/her judgement and understanding of interaction, but social content is *delivered* through a reduced channel of information. Virtual characters are of great interest to ASC-tech because they can provide simpler social information (e.g. using a limited, somewhat exaggerated set of facial expressions), and may be less demanding or confusing to interact with than a real person. Interaction with a child-like virtual character may have a very different dynamic than an adult character, or direct interaction with an adult: in school or at home, adults are authority figures and may be "in charge" of most interactions. A situation with an adult may be automatically understood as one where there will be demands, and right or wrong actions. A peer may not have these associations.

A virtual peer with a limited set of language and actions, controlled via Wizard-of-

[27] Unfortunately, these strategies are not described in great detail. We are not given a script for Sam, or any of the interaction guidance or goals used by the controlling researcher. It is also not completely clear from the paper whether TD children received any additional instructions, such as to invite the other child into the interaction/story as much as possible.

Oz, appears to marry some of the advantages of people and technology when it comes to this type of play-like situation. This study, and related work on virtual characters or peers for autism, are a clear illustration of how technologies may bring new affordances and opportunities to autism support and intervention, that are not available through paper-and-pencil methods. It also illustrates how technology can go beyond specific activities or skills practice exercise, to target communication through interesting, interactive *situations*.

Tablet apps for fostering social communication between children with ASC Hourcade et al. (2011) use a totally different medium (tablet computers) but pursue a broadly similar strategy to that of Tartaro and Cassell (2008). They developed a set of four multi-touch tablet PC applications, in collaboration with practitioners and parents. The goal of the application suite was to explore how collaborative activities on tablets might foster children's social engagement, communication, and self-expression. This study was not an intervention, and did not aim to directly teach skills. Instead, it again set up interesting, interactive *situations* (now in the form of games or apps) that tried to "tempt" communication (Wetherby and Prutting, 1984). These form the basis of shared, social interaction that could be scaffolded by an adult facilitator and (potentially) peers.

The three applications with the greatest social communication component (as described in the 2011 paper) were as follows. All were used in small lunchtime sessions that included children with ASC and typically developing classmates, with an adult facilitator[28].

Drawing A free-hand, zoomable and rotatable drawing application (using a stylus). Children used this individually, and also as part of collaborative storytelling. One child would begin a story, supported by a drawing, and then would need to pass the tablet to the next child who would add to both of these. The authors note that "both children with ASD and typically developing children show great interest in how the story will turn out, asking their peers about what they are drawing and what will happen next" (Hourcade et al., 2011, p.161).

Music authoring Children select tiles from a grid-like display, each representing a different note, and then their 'composition' is played column by column across

[28]"Lunch bunches", an established and familiar activity at the school where the evaluation took place, rather than a situation set up for the purposes of the research.

the screen. Like drawing, this application was also tried with collaborative authoring, and children showed similar enjoyment and interest.

Untangle (visual puzzles) The authors explain "The puzzle consists of a set of circles, each connected to 2 other circles by lines. To solve the puzzle, none of the lines should overlap" (p.161). Children must solve it by moving circles, "untangling" the lines. This quickly requires children to negotiate about their moves, because any one move changes the overall configuration. The game can require greater collaboration by using two circle colours, and assigning each child to move only his colour. The authors note that children sometimes began to spontaneously make suggestions to one another about moves to take.

For all three applications, the authors note that the adult facilitator's support was required to help children become acquainted with the technology, and manage the overall interaction (e.g. allowing another child to have a turn). They report in detail about how several children with ASC interacted with the applications, peers, and adults over time. What is notable from their descriptions is both the variability of how children interacted and what they seemed to find engaging, but also how the technology was a useful object of, and structure for, shared attention and communication between partners. The authors comment on these mediating and structuring functions directly, noting "the computer itself becomes the recipient of the participants' focus, providing a more structured narrative for the interaction patterns, rather than the usual open-endedness of social interaction, which perhaps reduces the anxiety of initiating social interaction" (p. 166). Speculating on these results, the authors suggest that the children's mutual interest in computers (for both ASC and TD), combined with applications that tried to support social skills, might have served to make their social interactions inherently rewarding. In a key qualitative result, when participating children gave feedback on what they liked about the activities, the *social* part of the activities was repeatedly mentioned, not the technology alone (see p. 166).

The applications were evaluated in more depth in Hourcade et al. (2013). Notable findings include that children spoke more sentences each minute and made more verbal interactions overall while using the apps than in a equivalent non-app activities (e.g. drawing on paper). The music and Untangle apps were also particularly successful at encouraging supportive comments[29] between children, compared to the other apps and no-app conditions.

[29]The authors included verbal support, encouragement, and helpful suggestions in this category.

General commentary on Tartaro and Cassell (2008) and Hourcade et al. (2011) Both these examples are organised around a game, or playful situation that provides opportunities for learning and skills practice. Many other ASC-tech examples with a variety of underlying approaches and physical hardware also are overtly or implicitly game-like, as a route to education or behavioural intervention (e.g. Dautenhahn et al., 2004; Fletcher-Watson et al., 2015; Malinverni et al., 2016; Gotsis et al., 2010; Tanaka et al., 2010; Whalen et al., 2006a; Wainer et al., 2014)[30]. The literature suggests that turning things into games is frequently a motivational choice, to make content or activities more engaging and more entertaining for children. Young children *may* find games more fun than unadorned drill-and-practice tasks—though this does not necessarily mean they see games as un-work-like, especially when they perceive directions or expectations from adults (Wing, 1995). Importantly for children with ASC, games have some kind of rules and/or goals. This can provide a necessary degree of structure; it may be much harder for them to engage in a completely open environment.

As a final point, note that both of these detailed examples are at a proof-of-concept scale. They are trying to determine the potential value of an approach, before revising it and perhaps scaling up its goals and participants. Hourcade et al. (2011), which involved parents and children in the development process and was evaluated during an established, inclusive activity time at the school, take great pains to describe the school environment and the children. This is a necessary context for understanding what they have done, and for other researchers to judge the relevance of their work to other situations. While Tartaro and Cassell (2008) give less context, they do still report many example pieces of dialogue to illustrate what the interactions were like. As with these examples, the majority of other work in ASC tech is small, local, and particular, with great attention to context and child characteristics (see Chapter 3).

2.6 Placing initial ECHOES observations in the context of the literature

The type of interactions described in the thesis introduction, with children reacting interestedly and positively to unexpected changes and rule-violations in their environment, are very surprising in light of the autistic "need for sameness". This is not just an observed characteristic of the autism spectrum, but a *core feature*, part of the diag-

[30]These projects target a range of skills and issues, not only communication.

nostic criteria. The disparity between this characteristic and the phenomenon observed in ECHOES raises an obvious question: what is going on here? Given the list of purported benefits of technologies, particularly VEs, as controllable, predictable, repeatable spaces, the current observations seem doubly ironic. Two decades of papers have repeated the assertion that computers are *particularly* suited to supporting and teaching users with ASC because of their capacity to provide sameness, repetition, and routine. ECHOES has done well at (fortuitously) motivating communication when the system introduced new things, got stuck, violated patterns, and so forth. Offering an explanation for this apparent conflict between the need for sameness and the current observations becomes a key task of the research.

The type of child interactions observed as a part of this new phenomenon are also unusual. It is unexpected but extremely desirable to find children with ASC spontaneously initiating communication, especially as some of these appear to be socially motivated. Initiation is developmentally important and there is a clear incentive to help children acquire and use these behaviours if possible. However, it is difficult to directly target initiation skills in a study or intervention programme, due to the requirement for spontaneity. As soon as a partner tries to prompt or elicit a child communication, the child's action becomes a response to the partner, not a "new" communication. One strategy has been to promote initiation through indirect means, creating demands or cues in the environment, especially as a part of familiar activities. The circumstances of the spontaneous child initiations in ECHOES suggest a link to these environmental modification strategies. They are not a direct precedent or foundation for the current work, first of all because they concern face-to-face contexts. The existing sources are also often guidance-only, descriptions of single children drawn from clinical practice, or strategies for eliciting behaviour in an assessment context. The one experimental study that could be located in this area, McClenny et al. (1992) does not include participants with ASC. Approaching from the angle of autism and technology, this thesis may be able to provide experimental data to further illuminate this general concept, and offer some comparison between different modifications, children, and communications.

The current observations in ECHOES suggest a line of investigation that can lead to a *potential new design strategy for autism-targeted technologies*, and a means by which to encourage initiation behaviour. Given the difficulty of targeting initiation and its far-reaching developmental importance, any new or expanded strategies in this area are potentially very valuable—especially if children find the experience to be positive, enjoyable, or humorous. Creating positive affect and a sense of reward around social communication is of benefit in itself, especially when interaction is so often cast as being stressful and demanding for children on the spectrum.

2.7 Summary

This chapter described the key diagnostic criteria of the *autism spectrum conditions* (ASC), according to the APA (2013). It discussed each main area of difficulty in more detail (social communication, restricted and repetitive behaviours), commenting on how these may affect people with ASC in their daily lives and over the life span. The remainder of the literature review focused on social communication, particularly initiation.

To better understand the idea of initiation and how it differs from responding to another person's communication, this chapter introduced the distinction between prompted and spontaneous communications, or, those with or without a clear antecedent in a partner's behaviour. With this distinction in mind, the review considered some of the specific difficulties that people with ASC may have with initiation, and how these can impact longer-term life outcomes. There was a brief consideration of some main approaches to social communication interventions, from highly-structured and adult-led approaches, through transactional-type approaches, to those that prioritise positive child-adult relationships and entering into a child's world. Strategies in which a child's environment or routines are altered to promote communication use (i.e. communicative demands, communicative temptations, staged problem-solving situations) were discussed at some length, as this literature is quite fragmented and in some cases difficult to access. Various authors present similar suggestions for altering familiar situations to demand or motivate child communication in some way—such as placing a favoured object in a visible but out-of-reach place. One only study, McClenny et al. (1992), appears to have empirically investigated the effectiveness of these situations at motivating child communication. The other sources, frequently for parents or practitioners, mention anecdotal data where they give any concrete examples at all. This small body of work suggests an indirect precedent for the investigations in this thesis, and also highlights that more experimental evidence for this area would be a valuable contribution.

A brief review of technologies targeted at people with autism (ASC-tech) focused on the rationale given in the literature for *why* technologies might provide unique benefits and affordances for autism. Two specific examples of technologies that target social communication were discussed in depth, as an illustration of how different projects have tried to indirectly support and scaffold communication around technology, similar to the broad approach in the present project. Finally, an overall discussion section pulled out key lessons from the current literature, and used these to frame the initial child interactions reported in the thesis introduction and highlight where the current investigations may fill gaps in existing work on spontaneous communication and ASC-tech.

The following chapter sketches out the research approach for this thesis, and includes additional methodology literature and overview of ASC-tech as a research field. It will also return to the research questions introduced in the introduction, identifying sub-questions and the range of methodologies needed to pursue each part of the investigation.

Chapter 3

Research approach

3.1 Introduction

This chapter seeks to lay out the overarching logic and priorities that guided methodological choices throughout the work, and place them in the context of the autism and technology sub-field. The main determinant of methodological choices was, of course, the need to answer the research questions. As already identified in the introduction, this thesis represents a unified line of investigation pursued through successive phases of research, requiring different types of activity. Each phase built on the previous phase, with its specific questions developed in reference to past results. Given the novel, poorly-understood observations that initially suggested this research, it was not possible to plan a single set of questions to guide an overall programme of work. However, the final questions can be presented as a sequence which encapsulates the complete investigation. These questions, listed in the introduction, are expanded here among with their sub-questions, and connected to the methods used to answer them.

While a complete set of questions could not be set early in the research process, boundaries for the work *were* defined early on, constraining and prioritising later choices. For example, it was an early decision to investigate the emerging concept of discrepancy with an emphasis on its relationship to, and implications for, children with autism—as opposed to seeing it as an interesting phenomenon for interaction design more generally, in which children with autism are an example group of users.

With these priorities in mind, this chapter identifies the overall approach of the research as pragmatic, and with "mixed methods". It blends qualitative inquiry with strategic use of numbers to illustrate and summarise a novel, complex dataset. Following from this overall positioning and the current practices of ASC-tech as a sub-field, *transferability* and construction of *working hypotheses* (Cronbach, 1975) become key goals of this work. These concepts are explained in some detail, as they are particularly important to understanding what the latter part of the thesis (Chapters 6-8) does and does not seek to accomplish.

Details of specific methodologies are reserved to the chapters in which they are applied. Those chapters give further justification for specific methodological choices, within the priorities and limits outlined in this chapter.

3.2 Researching and writing technologies for ASC: Standards, evidence, and where work "lives"

3.2.1 ASC-tech as an emerging sub-field

Technologies specifically developed for, or applied to, the needs of people on the autism spectrum form an emerging sub-field often referred to as *ASC-tech*, *ASD-tech*, or *technologies for autism*. This thesis will refer to ASC-tech and technologies for autism interchangeably. As the term is used here, it does not refer to basic research that may *use* technology to study autism as a condition, such as eye-tracking or brain scanning research. It refers to technologies used directly by, with, or to support people with autism. For example, software that helps special education teachers to create visual supports for pupils with autism would be included in this definition.

Emerging is perhaps a misleading word for ASC-tech, because the area of inquiry is not new. Researchers have been uniting autism and technology for decades, operating from within their "home disciplines" of psychology, computer science, education, speech and language therapy, and design. Represented herein are an astonishing diversity of methodologies, from phenomenological inquiry to case studies to randomised control trials, and a range of views around the autism part of "technologies for autism"[1]. Within any of these "home disciplines" and regardless of specific methodology, ASC-tech is usually treated as exceptional in some way: technology is an unusual and novel basis for support or intervention, compared to existing practices in education and psychology. In human-computer interaction, people with autism are a challenging, special-case user group. Researchers and practitioners have thus far had

[1] For example, researchers' discussion of autism and the goals of their technologies can vary substantially, from a very traditional medical or "deficit" model, to more social or neurodiverse conceptions. Some researchers, especially those more interested in technology, may simply avoid this entire issue.

little choice but to repeatedly argue for ASC-tech as an area worthy of investigation.

Only recently has ASC-tech begun to coalesce as a distinct sub-field, writing and citing across disciplines. At the time of this thesis, there are still no journals focused on ASC-tech, and limited conference opportunities in which researchers can meet as a community. The lack of dedicated venues and the need to continually justify ASC-tech research has been a major barrier to the field moving on to more complex conversations that address the challenges of ASC-tech *as a field*, including effective and productive models of research[2]. Consequently, ASC-tech work is still almost exclusively conducted following the customary practices of its researchers' home disciplines, and prioritises the types of evidence and activity valued in those disciplines (or, more practically, those required for publication). What is considered "good practice" or "convincing evidence" for ASC-tech researchers operating out of one discipline would not pass muster for researchers in another discipline—a challenge common to other types of interdisciplinary research.

Calling attention to the diverse and currently fragmented state of ASC-tech is meant as an explanation for why there is no standard model of research in this area—and may never be, due to its fundamental interdisciplinarity. In place of a few standard (or at least frequently-employed) models of work, there are *many* models, determined by researchers' disciplinary backgrounds and the audience(s) which they hope to address. There are multiple types of knowledge production that are considered legitimate and valuable. In some ways, this diversity of evidence and knowledge production is actually a great strength of ASC-tech. In the absence of established ways of working and standards of evidence, it is particularly important for ASC-tech researchers to signpost their methodological commitment and priorities. What standards they are endeavouring to meet? This signposting makes it clearer what a piece of research does, and does not, try to do, and aids readers in judging the work by appropriate measures. This chapter, particularly Section 3.3, plays that role for this thesis. At the same time, it hopes to provide a model for working in a way tailored to the needs and concerns of

[2]For example, there has as yet been little progress on how to balance rigorous evaluations with the rapid speed of technological change. By the time evaluation of a technology-based intervention is completed, following traditional methods in education or psychology, the technology used in the intervention may be outdated, no longer supported, or supplanted by something more exciting that parents and schools have purchased instead. The technology may *certainly* be outdated by the time work is published! A quick-and-dirty evaluation may be of greater value and impact than a more careful one. There is as yet no consensus on what to do: should we move the goal posts on the type and quantity of evidence needed to say that a technology is effective? It may be extremely difficult for researchers to try out the options and identify workable models of *technology* evaluation, when their funding and publication options effectively demand other types of evidence.

ASC-tech as an emerging discipline.

3.2.2 Keeping things small and focusing on context

The vast majority of ASC-tech research that has developed and/or directly evaluated technologies[3] shares two things: it is small, and context is absolutely crucial to producing and understanding the work. Small-n designs are prevalent. Thirty participants is impressive, and 100 is almost unheard of. Why does ASC-tech research tend to happen on such a small scale? There are a number of intersecting factors, which could be summarised as follows:

Heterogeneity of autism Autism is a spectrum condition. While there are core diagnostic characteristics, individuals may be affected to different extents in different areas. People who look similar on paper in terms of standardised measures or diagnostic categories may be *extremely* different in terms of daily-life strengths and challenges. Researchers wishing to do controlled research in ASC-tech (or any autism research) with participants similar to one another face a trade-off: unless a study is exceptionally well-resourced and operating over a large geographic area or a long period of time, recruiting more participants will almost always require more flexible inclusion criteria. It will also expand the pool of *other* supports and interventions that children might be receiving, outside of a particular study.

Heterogeneity of contexts and attitudes It seems obvious, but is worth stating that the contexts in which people with autism may receive services and support (and the supports received) will vary substantially. Each school or family is a unique context, but as a group, schools or families in the UK will have different features and concerns than schools or families in the USA. As one example, these two nations have widely divergent attitudes and policies toward the extent to which children should be placed in mainstream versus specialist education. A technological tool designed in reference to one context may be substantially inappropriate in another.

Limited participant pool There are far fewer children with ASC than there are typically developing children. By the time that researchers narrow down to a target

[3] As opposed to work surveying technology use, or technology attitudes (e.g. Putnam and Chong, 2008).

age group or level of ability, the pool of qualifying children may be small indeed. In the UK and around the world, the autism community also regularly receives far more requests for research participation than can feasibly be accommodated. Researchers are frequently in competition with each other to secure *any* participants, let alone large samples.

Technology-related barriers Depending on the type of study, it may not be feasible for researchers or participants to provide (and support) large numbers of devices, especially where these are expensive, fragile, or custom-built for the research.

Appropriateness of exploratory research As yet, the corpus of ASC-tech research is small compared to more traditional face-to-face or paper-based methods for studying, supporting, or teaching people with autism. Moreover, available technologies constantly change. Exploratory, proof-of-concept-scale research is frequently the most appropriate choice for determining the potential impact or feasibility of new methods in this area. Work of this type *must* be done in order to intelligently scale up to larger field studies, or more controlled experimentation.

ASC-tech as applied, contextual research Thus far, technologies for autism have almost exclusively focused on providing concrete tools. The sub-field is about applied, rather than basic research. Context, usability, and goodness-of-fit with existing practices are all of *vital* importance. A beautifully designed tool can be scuppered by the fact that a school may not give teachers permission to install new software. Case study-like research is able to consider the interaction between designs and contexts in a way that may be difficult in larger-scale research.

While a small, heterogeneous participant pool is an undeniable barrier to doing larger scale ASC-tech work (or indeed many kinds of large-scale autism research), the final two reasons are the most important. Even when research on technologies for autism takes place in a lab, rather than a school, home, or practice setting, its end goals are about *applied work*, and there is currently limited information about the approaches, hardware, and implementation practices that will be effective. While the field may mature into a different way of working, with larger-scale or more controlled studies, at present it is still engaged in substantial "map-making" type work, looking for proof-of-principle for different technologies and methodological approaches.

While work in the lab can provide valuable, formative information, ASC-tech is intended for use in the world. Currently, smaller case study-like work provides the best information about how research technologies are working in context (or not), and their interactions with specific individuals. These many smaller-scale studies also begin to identify and investigate the practical challenges that must be solved in order for self-maintaining ASC-tech to become a reality in the future—that is to say, employed by autistic individuals, parents, teachers and service providers in business-as-usual, not as part of pre-planned, researcher-directed studies.

Large-scale work with ASC-tech can be very difficult in a practical sense, and is frequently premature given the current state of the field and the incessantly moving goal-posts in terms of available and desirable technologies. Beyond that, its current aims are actually better served by work that acknowledges the heterogeneity of individuals and contexts, and uses this as a tool for greater understanding. The work reported in this thesis is small-scale for all of these same reasons. It explores a novel, poorly understood phenomenon, to generate information for future work and to determine what type of future work might be meaningful and appropriate. Instead of a surface understanding of a larger sample, it seeks to deeply understand the behaviour of particular children, in a particular context.

A key purpose of this type of small-scale, contextualised work is, ultimately, to *transfer* useful design insights from one context to another. The onus is on the transferring researcher or practitioner to judge the "fittingness" (Lincoln and Guba, 1985) between an original research context and a new context. Even where individual technologies and contexts of use are unique, there will be similarities that enable a degree of transfer. The heterogeneity of autism and of ASC-tech largely preclude generalisation, because, to be meaningful and feasible, all ASC-tech work must be conducted in relation to particular contexts, and will produce local, provisional tools. Transferability is a much more relevant goal and measure of value for the ASC-tech sub-field than is generalisation.

3.2.3 Transferability

The introduction to this thesis stated a goal of creating tools to support *transfer* of the lessons learned in ECHOES to other related contexts. Similar to its everyday sense, transfer in research means taking something from one place or sense to another one. Successful transfer of findings requires that the original author in a *sending context*

(Lincoln and Guba, 1985) provide detailed, rich information about the particulars of a piece of qualitative research. The extensive context and examples provided (in the manner of *thick description,* Geertz, 1973) allow the reader to judge whether the account given is fair, trustworthy, and plausible. A reader with knowledge of this sending research context as well as his or her own potential *receiving context* can make a judgement about the extent to which the original findings relate to, or may be applied to, their new context (i.e. may be transferred). In short, transfer in qualitative research is about using a deep knowledge of a particular context to understand aspects of other contexts.

This goal of transferability is in contrast to generalizability, which is concerned with making statements (*generalizations*) about a subset of a population, which can be extended to the population as a whole. Generalizations "formulate what is always and everywhere the case, provided only that the appropriate conditions are satisfied" (Kaplan, 1964, p.91). They are explanations, constructed based on "a representative sample of contexts" (Lincoln and Guba, 1985, p.124), and are understood to be universal, not time-limited, and applicable to all instances of a certain kind, or all members of a population [of contexts]. Here, the researcher's job is to illustrate that the starting dataset is representative and universal, and to clearly delineate its membership conditions. The researcher must demonstrate that the constructed generalizations are logical and plausible, following from the data.

As already indicated, generalisations are not particularly useful in the autism and technology subfield (ASC-tech), given its nature and goals. ASC-tech has (thus far) had little emphasis on producing generalisations, or on conducting tightly controlled experimental research[4]. Potential transferability is by far a more relevant measure of value in ASC-tech research than is generalisability. Even when each piece of research occurs in a unique context, there will be other projects that show related circumstances and concerns (Shenton, 2004). Where there appears to be good fit between existing work and potential new contexts, transfer of specific techniques and findings is certainly possible.

Providing support for transferability through reporting "thick description" and "lessons learned" is common within ASC-tech, although authors rarely signpost these actions in terms of transferability. They may *imply* transfer, and refer to other authors borrowing from or building on their findings. For example, in reporting their autism design guidelines from the LINKX project, van Rijn and Stappers say explicitly that they "want to

[4]There are a few counterexamples: see Fletcher-Watson et al. (2015) for an example of a Randomised Control Trial (RCT) using an iPad app for autism.

share [their] guidelines...so that others can use them as stepping-stones in their work" (2008, p.2). Other authors, not writing direct design guidance, still frequently provide the detailed contextual and interactional information necessary for other researchers to judge "fittingness" between contexts, but without ever overtly mentioning transferability as a goal of their work (e.g. Parsons et al., 2004; Keay-Bright and Howarth, 2012; Hourcade et al., 2011).

A qualitative, situated project may still produce explanations more general than that of idiographic knowledge, though these explanations will be different than the universal generalizations produced in a tightly controlled, experimental project. What we may produce are *working hypotheses* (Cronbach, 1975), or provisional, local, relative explanations. The working hypothesis is a sister to Stake's *naturalistic generalisation* (Stake, 1978),[5] which is based on recognising two kinds of generalisation: one nomic, universal, abstract, the other based on direct experience, and more akin to the type of understanding we develop about everyday life. Whichever term we choose, both represent explanatory concepts that *do* generalise and abstract, but "give proper weight to local conditions" (Cronbach, 1975). They do not conclude, but rather *speculate* on the potential extrapolation of current insights to other situations with similar circumstances (Patton, 2002).

As a group, the research questions presented in this chapter and the work that answers them later in the thesis are all steps on a path to transferability. The insights from an initial case (ECHOES), or *sending context*, are presented through a mixture of detailed description and quantification, to communicate both the sense of "what happened" and the overall composition and extent of DR pair episodes in the dataset. The result are of course the results of an entire, holistic situation. Rather than entirely deferring "fittingness" and relevance judgements to future researchers, a further spell of reflection sought to speculate about their application to a new situation, with similar circumstances. The production of design guidance (Chapter 6) was intended to highlight, based on a deep knowledge of the initial ECHOES concept and its development over time, *which* contextual features would need to be similar to the receiving context in order to facilitate transfer. The final part of the thesis (Chapters 7-8) puts this transfer into practice, and evaluates the extent to which it appears to have been successful.

[5] Some authors treat them as effectively equivalent (e.g. Lincoln and Guba, 1985; Lewis and Ritchie, 2003). Here, the term *working hypothesis* will be used throughout the remainder of the thesis.

3.3 Prioritising ASC within “designing technologies for ASC”

3.3.1 Understanding and applying discrepancy in relation to ASC

Moving from initial observations of discrepancy to more systematic analyses[6] posed an important choice regarding the focus of the current work. One option was to focus on how the design and affordances of technology may meet the interactive, social, and cognitive needs of children with ASC. Another option was to consider discrepancy and expectation-violation as a more general phenomenon or strategy in human-computer interaction. Children with autism, or more specifically their communication difficulties, would then be an example group of users to which the potential strategy was applied. Alternately, the work could have sought to make connections even further afield within computing or education, for example linking it to research on erroneous examples as a pedagogical strategy in intelligent tutoring systems (e.g. Adams and McLaren, 2013; McLaren et al., 2012).

This work takes the first route, continuing to prioritise autism. The goal is to understand discrepancy and its potential applications in technologies for children with autism, rather than seeking to understand discrepancy and technology generally, with autism as a particular use case. Children with autism are a distinct user group with respect to interaction design. Designing for this group differs from designing for “young children” generally due to two central issues: the characteristics of autism, and the end goals designers may be trying to achieve. The core autistic features of difficulty with communication and social interaction, and the presence of restricted and repetitive behaviours (RRB)[7], drive the divergence from common design goals for young TD children. A sizeable proportion of all ASC-specific technologies target skills that TD children may acquire early in life with little or no explicit instruction[8]. Thus, there is a need for design strategies that engage with both ASC-specific characteristics and the social and pedagogical goals relevant to this group. Pursuing those goals and meeting the needs of people in this group (e.g. for comprehensible and manageable interactions with technology) may *not* result in innovative design or technical novelty. It also may not result in insights, methods, or products that are applicable to other user groups or

[6] A process reported in Chapter 4

[7] Discussed in depth in the literature review (Chapter 2).

[8] Such as recognition of facial emotions (e.g. Baron-Cohen et al., 2009), or preferentially orienting to people (Fletcher-Watson et al., 2015).

types of problems. Some or all of these *may* result from the current work, but they are not a requirement for its success.

The focus on understanding discrepancy in relation to autism and designing for autism-specific communication challenges means that control groups of typically developing children (or children with other disabilities) are not a useful tool for the current work. While TD children helped to test the new game designs (Chapter 7), *their participation in the games evaluation would not help to answer the current research questions that are exclusively focused on autism.* The literature is very clear that children with autism struggle with initiation and other communication skills that TD peers learn easily and without explicit instruction. TD children do not need specific design strategies or interventions to support this skill set. Comparing the two groups of children would almost certainly reproduce patterns well-known from other studies—with TD children initiating more often, with a greater range of behaviour and purposes—and would not help us better understand the needs, interests, and interactions of children with ASC.

3.3.2 Design and ASC

The current project is about investigating technology *design* strategies relevant for children with ASC. Once past the initial issue of understanding and describing discrepancy as a phenomenon, the work is interested in the interaction design potential of discrepant aspects. The focus on design in some parts of the thesis has implications for methodology and reporting, because (in general) design will have different ways of demonstrating success and accountability. Gaver (2014) suggests that the basic contrast between accountability questions in science and design is about processes of knowledge production, versus workability. Questions that we ask about a piece of scientific work are "variations on a single one: 'how do you know what you say is true?"' (Gaver, 2014, p.146). Design, he says, "works with aesthetic accountability , where 'aesthetic' refers to how satisfactory the composition of multiple design features are (as opposed to how 'beautiful' it might be). The requirement here is to be able to explain and defend—or, more typically, to demonstrate—that one's design works" (Gaver, 2014, p.147). Asking about epistemology versus workability means that science and design work are subject to different expectations, particularly at the point of being explained to or evaluated by an audience. Those two types of questions need different information to answer them, and a "legitimate" and convincing account of the work will look quite different.

A single piece of work may certainly combine both science and design. When it does, it must also prepare to give a combined, or perhaps *dual* account. Different parts of the work may need to ask different kinds of questions about knowledge production versus workability, and give different accounts of what the work is, knows, or does. This thesis does combine both scientific and designerly activity, because both are needed to meet its set of goals. It seeks to coherently narrate "how it knows" and "how it works" at different points. As Gaver advises for projects with "intertwined" science and design, it seeks to "[be] clear about the form of accountability claimed for different aspects of the process and results" (2014, p.163).

The later parts of the thesis (RQ4, RQ5; Chapters 6-8) focus on developing, implementing, and reflecting on a new design. The goal is to understand and assess the effects of that design (or, its workability). Children are playing the role of *testers*, per Druin's (2002) discussion of possible child roles in technology design. The researcher observes children interacting with a technology, to gain insight about the usability and effects of the design. As emphasised elsewhere, the work in this thesis is not an intervention that seeks to teach children new behaviours or increase their incidence of certain behaviours. It is only interested in what happens *while* children are actually using a design. *The evaluation study is an evaluation of how the design works*, and in no way an evaluation of the participating children and their skills. For this reason, it also does not compare individual children in terms of "performance", but only compares them to illustrate the range and variability in child-technology interactions.

3.4 Overall methodology and research questions

3.4.1 Pragmatism and fitting techniques to questions

The current work takes a pragmatic approach to methodologies. It treats qualitative and quantitative techniques *as a set of tools*, rather than as opposing viewpoints or sets of assumptions about studying the social world. In other words, it espouses pragmatism, "focusing on research methods as techniques divorced from their philosophical foundations" (Snape and Spencer, 2003, p.18). Snape and Spencer give a fairly typical view when they point to goodness-of-fit between methods and research questions as an overriding concern, even if this means choosing a set of methods that may have dissimilar underlying epistemologies (p. 21). They are clear that they do *not* think this undermines research quality. Indeed, Bryman identifies this fit, or what he calls "dovetailing" of techniques and research questions, as being *the* prime concern when researchers treat methods as tools, rather than philosophical commitments (Bryman, 1988, p.5). It is this concern for good fit and tight dovetailing that drives the methodological choices in the the current work.

A broadly qualitative approach is the most appropriate for the current series of inquiries because, as noted in the thesis introduction and literature review, this is a novel phenomenon with little related literature. The less-structured, generative, and contextualist[9] nature of most qualitative methods are well-suited to research where there is little existing theory to guide a more targeted inquiry (qualitative or not). Qualitative methods perform well where the goals are understanding and description of (as yet) poorly-understood phenomena, and generation of new theories and strategies—what Ritchie calls the *contextual* and *generative* functions of qualitative research (Ritchie, 2003). Finally, qualitative methods can provide a particularly good fit for work in ASC-tech, because of the huge importance of context and values in technology design and evaluation.

3.4.2 Mixed methods

The overall approach taken across the thesis might best be described as "mixed methods", using this term as a general description, rather than a reference to a specific model or research design. It used both qualitative and quantitative data and methods to carry out a unified programme of research. Qualitative methods and modes of reporting were used in order to inductively build categories and concepts from data, understand the particularities of interaction-in-context, and provide "thick" descriptions of new phenomena. Numbers were reported in a primarily descriptive way, to communicate the prevalence, extent, and typicality of the phenomena described. Statistics were used to calculate inter-rater agreement of data annotation, and whether apparent differences between categories are "real" differences, in a statistically significant sense.

Counts of category instances and simple statistics, as in this research, can help to extend and validate descriptive accounts of the phenomenon under study. These give evidence about the representativeness and/or typicality of those accounts. Otherwise, authors are open to the "charge of anecdotalism", or that they have cherry-picked events and selectively presented descriptions in order to support their argument (Sil-

[9]Having a "commitment to understanding events, behaviour, etc. in their context" (Bryman, 1988, p.64).

verman, 2006, p.298). These may in fact be atypical, and unrepresentative of the data as a whole. Using counts to give an overall picture of a complex, varied dataset can counter these charges and convince the reader that the account is representative of the broader data set.

3.4.3 Research questions and methodologies

The thesis introduction briefly explained how this work has been conducted in stages, with each stage posing questions and developing working hypotheses that guided subsequent stages. It is not possible to state a set of overall questions that guided the entire research process. Instead, a sequence of high-level questions captures the *line of investigation* followed through the thesis. The five main questions are expanded here from the version in Chapter 1, and presented with their sub-questions. For clearer mapping to the thesis content, these questions refer to discrepancy throughout, even though this was a concept that was developed partway through the work and underwent substantial revision.

RQ1 What is happening when children with ASC were observed to communicate about novel and apparently surprising aspects of ECHOES? What is the nature of this phenomenon (eventually labelled as *discrepancy*)?

RQ2 Is discrepancy represented by a few rare but striking instances in the ECHOES dataset, or are there sufficiently widespread examples to merit further investigation and form a basis for categorisation?

RQ3 Given that instances of discrepancy appear repeatedly across ECHOES sessions and participants, how can we describe and categorise them? Which features or sources of information need to be included in order to capture the nature of these instances?

- a In what ways do children react to *discrepant aspects*?
- b What things or events in the environment appear to be a source or cause of discrepancy (i.e. discrepant aspects)?

The first three questions are mainly addressed in Chapter 4. Early phases of analysis provide answers to RQ1 and RQ2, establishing a sense of “what is happening” with child reactions to discrepancy, and that this phenomenon appears repeatedly across the dataset. Chapter 4 reports the process of moving from initial observations of discrepancy in ECHOES video data, through various types of iterative labelling and categorisation activity (with particular reference to Grounded Theory) in order to develop taxonomies of discrepancy and child reaction, based on ECHOES data (RQ3). It lays out the evolving reasoning leading to the final "core category" of discrepancy, which encapsulates the nature of this phenomenon.

Chapter 5 systematically applies the newly-derived categories to the ECHOES dataset, to produce descriptions and counts of the ways in children are reacting to discrepancy (RQ3-a). This includes considering children's affect in relation to discrepancy. Chapter 6 conducts a separate process of successively labelling and grouping environmental aspects into categories, to answer RQ3-b. As with child reactions, category counts summarise the complexity of the dataset beyond the examples described in detail, and communicating the notable *lack* of "typical" instances.

Working through these first three questions allowed the final questions to be set, as a pair: RQ4 asks the questions necessary to design and implement new technology that tries to leverage discrepancy, and RQ5 then asks Gaver's (2014) central accountability question "does it work?".

RQ4 Is it possible to create discrepancy and discrepancy-contingent communications on purpose? How might this be achieved?

- a Per the concept of discrepancy as a *relation* or *state of affairs* between the child and the environment, how would a designer try to facilitate this phenomenon?
- b What features of the ECHOES context appear to have contributed to a *motivating but manageable* experience of discrepancy?

RQ5 Guided by design principles rooted in empirical work with ECHOES and autism theory, can designed *discrepancy-detection opportunities* (DDOs) successfully create a *motivating but manageable* experience with new participants, in a new set of games?

- a Will children be motivated to spontaneously initiate social communication about the designed DDOs?
- b Will children's game-play experience be emotionally manageable, per their affect, utterances, and behaviour during games sessions?

c How will specific design decisions affect children's communication and their overall game-play experience?

This pair of questions is about transferability. The data and analyses produced in answer to RQ1-RQ3 give a detailed picture of the *sending context*, and Chapter 6 prepares to apply its insights by working out what it would actually mean to create discrepancy in practice (RQ4-1), extrapolating from the concepts developed earlier. It also reflects on the features of ECHOES that appear to have contributed to positive, discrepancy-contingent initiations, formulating these into high- and low-level design guidance (RQ4-2). Chapter 7 then designs and implements a custom *receiving context* in the form of new games that will be evaluated with children on the autism spectrum, producing data with which to answer RQ5. Chapter 8 reports the results of children playing with the new games, which are used to give an account of how the design "works" (RQ5-a, b). Reflection on RQ5-c takes place, at different levels of abstraction, in both Chapter 8 and the thesis general discussion (Chapter 9).

3.5 Summary

This chapter outlines the research approach for this thesis, and places its choices and commitments in relation to autism and technology (ASC-tech) as an emerging sub-field. This field is as yet fragmented, with researchers contributing from their "home disciplines", and evaluating success and value according to the standards of those disciplines. There are as yet no shared standards or methodologies that characterise work in ASC-tech. This is not to say that work in the field does not have any shared issues and concerns—quite the opposite. ASC-tech work is largely applied, and most often small-scale, with a close attention to context. Reasons for this include the heterogeneity of the autism spectrum and autism contexts, a limited pool of participants subject to high demands for research access, and difficulty with accessing and maintaining technologies. Most importantly, exploratory and highly contextual research is currently what is needed in the field. There is as yet limited information about the approaches, hardware, and implementation practices that will be effective to meet various autism-related goals, and to accommodate sub-groups of users. At present, ASC-tech researchers are still engaged in substantial "map-making" type work, seeking proof-of-principle for different technologies and methodological approaches. The current state of the sub field explains why, at present, it is most useful for researchers to seek transfer of insights between related contexts and projects, rather than generalisations. The current project explicitly seeks to support transferability through creation of design guidance.

Even though it declares itself to work within ASC-tech, it is important for the research to spell out its specific priorities. It was an early decision to prioritise autism within ASC-tech, seeking to understand and meet the needs of young children with ASC *through technology*, rather than to understand the capabilities of a new technology *through children with ASC*, as a case-study. Children with ASC are a distinct user group in HCI, with distinct needs that are frequently *not* relevant to typically developing children or those with other types of disabilities. As the work investigates a novel phenomenon, discrepancy, it focuses specifically on how this is experienced by children with autism, and how it might be useful in autism-targeted design. This commitment to prioritising autism and the unique needs of this group are the reasons why there are no comparisons to typically developing children during the research: including them would not help to answer research questions about children with autism interacting with technologies.

The chapter proceeds to consider the overall methodological choices and positioning of the thesis. The work takes a firmly pragmatic approach, prioritising the appropriateness of techniques to answer particular research questions and considering their epistemological underpinnings minimally if at all. In taking this position, it joins a large body of existing research. The best fit for the research questions has been to mix qualitative and quantitative methods at different points of the investigation. These provide the best opportunity to explore and describe a poorly-understood phenomenon (discrepancy) and to generate hypotheses.

The chapter concludes by returning to the research questions introduced in Chapter 1, laying out their sub-questions and mapping these to the methodological positions and tasks introduced earlier in the chapter. These research questions together represent a line of inquiry with sequential stages of work, rather than an overall set of questions that guided the entire research process. The results of each research question were used to formulate the next questions. Later chapters in the thesis also return to methodology, discussing specific decisions and introducing relevant literature as needed.

Chapter 4

Developing a discrepancy-reaction taxonomy and video annotation scheme

4.1 Introduction

This chapter begins by giving further background on the ECHOES project and its summative evaluation, which provided the data used as the basis for iterative, inductive taxonomy development. Before stepping through details of taxonomy and annotation scheme development, it presents the core category of *discrepancy*, and the concepts of *child reactions*, and the *discrepancy-reaction pair*. This gives the reader the end point of the taxonomy development, toward which the following sections progress.

The analyses described in this chapter represent a messy, roughly cyclical process. The final result was two linked taxonomies of discrepancy and child reaction, as concretely implemented in the annotation scheme labels and their application rules. The taxonomy and annotation development is presented as five phases. The specific methodological steps and conceptual developments are discussed in terms of these steps. This chapter does not report all activity in the same level of detail, but focuses on key decisions and methodological choices that propelled the taxonomy/annotation forward.

The chapter concludes with a summary of the discrepancy-reaction annotation scheme in its current form, as it was applied in studies 1 and 2 (Chapters 5 and 8). This includes key definitions, diagrams showing how to apply concepts during annotation, and a worked example of how annotation labels would be applied to a codeable instance of video. The full annotation manual, with additional guidance and worked examples, is reproduced in Appendix C.

4.2 ECHOES project background

The ECHOES project developed a technology-enhanced learning environment targeted primarily at young children with ASC, but with the potential for use by TD children (Porayska-Pomsta et al., 2012; Foster et al., 2010; Alcorn et al., 2011). ECHOES included a programme of game-like learning activities set in a “Magic Garden”. It blended practice of foundational social skills (e.g. gaze- and point-following, turn-taking) with exploratory, playful elements intended to generally facilitate interaction. By bringing together what was (at the time) cutting-edge hardware and AI planning, with educational and psychological theory, the project team’s goal was to deliver an engaging learning experience suitable for children with a range of developmental trajectories. Participatory design (PD) was central to the process of developing ECHOES and making it appropriate, comprehensible, and meaningful for its users. Various PD activities within the project included children with and without autism, teachers, and other stakeholders (Frauenberger et al., 2011, 2012b,a, 2013).

4.2.1 Hardware, Magic Garden VE, and VC

The ECHOES VE was designed for a free-standing 42” multi-touch screen (see Figures in Chapter 1). A young child user standing or sitting in front of the screen could be immersed in the visuals and sounds of the Magic Garden, and be quite physically involved in the interactions (variously dragging, tapping, shaking, and tickling digital objects). Speech (child voice) and sound output were present, but were pre-recorded. There was no text-to-speech capability, meaning that there was no capacity to respond flexibly to children during sessions. There was also no capacity for speech recognition or other sound input.

The Magic Garden was home to Andy, an autonomous, childlike *virtual character* (VC), who was meant to be the child’s guide and playmate throughout the learning activities. The underlying AI software modules planned Andy’s behaviour both deliberatively and in reaction to the child’s actions (or non-actions) in the system (for details, see Porayska-Pomsta et al., 2012; Foster et al., 2010; Bernardini and Porayska-Pomsta, 2013). A researcher at a small monitor alongside the main ECHOES screen used a GUI to manage inter-activity transitions and give limited system commands (e.g. for Andy to repeat an instruction). ECHOES was not a Wizard-of-Oz system; the researcher's degree of control was limited, and mostly meant to keep the overall session flowing smoothly.

ECHOES activities all followed the theme of the Magic Garden and used related objects such as flower pots. The activities were designed to encourage experimentation and play by deliberately introducing novel elements and behavioural fantasy (the "magic" of the Magic Garden). Examples include "pulling" on flower heads to transform them into bubbles or bouncy balls, or a box of balls becoming a fireworks display when a sorting task is complete. The various ECHOES activities relevant for the discrepancy-reaction analysis are described in more detail in Appendix A.

4.2.2 In-school Summative Evaluation

Summative evaluation of ECHOES took place in Spring and Summer 2011 (results in preparation), and included 28 children with ASC from four UK school sites (see Chapter 5). The broad goal of the ECHOES summative evaluation study was to assess a variety of social and communication skills before, during, and after six to eight weeks of using the ECHOES environment.[1] Children completed several 10-20 minute sessions of game-like learning activities per week, with more complex material gradually being introduced. Video data was the primary record of the child's communication and social behaviour, as automatic logging captured touch-screen actions only. A subset of this video data forms the basis for the preliminary analysis reported in Chapter 5.

As an ECHOES foundation study discovered, children frequently interacted with the researcher(s) as well as the system (Alcorn, 2010; Alcorn et al., 2011). As a result, the camera was positioned to capture as much as possible of the broader study environment (screen, child, and researcher). Children sat on a chair within easy reaching distance of the touch screen. A researcher was always present during ECHOES sessions, stationed nearby the child to offer support, though the exact set-up and manner of researcher interaction varied between school sites. See Chapter 5 for a fuller discussion of researcher-child interactions, illustrated with images.

[1] See details in Chapter 5.

4.2.3 Researcher observations

The main ECHOES data analysis was a detailed behavioural annotation[2] of video collected during technology sessions and in structured and unstructured school contexts. The author played a main role in iteratively developing, testing, and documenting the ECHOES annotation scheme, and was a main annotator for session videos from several of the school sites. This was the opportunity for the original, unexpected observations of discrepancy, some of which have already been described in Chapter 1 (e.g. see Figures 1.1, 1.2). It was also an opportunity for immersion in the data, and for what Saldaña calls the Pre-coding and Preliminary Jotting stages of qualitative analysis, which identify particularly striking moments in the data and generate "ideas for analytic consideration while the study progresses" (Saldaña, 2009, p.16). Many of the analytic ideas in the final taxonomy and studies can trace their evolution directly from these earliest "jottings". The first phases of the current project sought to follow up on and explore these ideas in a systematic way, and to go beyond the most striking examples to see if there were others that might belong to a similar pattern.

4.2.4 Three main concerns of the analysis: child, environment, relations

The initial observations (briefly described in Chapter 1) identified what can be considered the three main concerns of the analysis that have persisted throughout the taxonomy and annotation development: the child's experience and/or behaviours, the environment and its events/aspects, and the relationship(s) between these two. Different phases had different ideas about the nature of each concern and their relative importance. Later sections resume a phase-by-phase discussion of taxonomy and annotation development, considering them along these dimensions.

4.3 Discrepancy and child reactions

4.3.1 Discrepancy as subjective inconsistency

Discrepancy, as initially defined in Alcorn et al. (2013b), is a high-level conceptual category referring to situations in which the child considers an aspect of the virtual

[2] The annotation scheme was an adapted version of the SAP, or *SCERTS Assessment Process*, a tool in the SCERTS framework for autism education and support (Prizant et al., 2005, Volume 1, Chapter 7).

environment to be somehow inconsistent with his/her current knowledge of, or expectations about, the virtual environment[3] and its contents.[4] In order for a discrepancy to occur, a child must have begun to build up knowledge and expectations through interacting with or observing the environment. This body of knowledge and expectations will continue to expand and evolve over time (Figure 4.1, a). Now possessing knowledge and expectations, s/he *may* identify subsequent interactions with aspects of the environment[5] as being consistent—or inconsistent—with these expectations. In the majority of interactions, the child is likely to find the environment consistent with its own past behaviour, and with his or her current expectations; this situation is termed *expectation fulfilment* (Figure 4.1, b). However, in some interactions, the child may consider an aspect of the environment to be somehow *inconsistent* with his or her current knowledge or expectations. This situation, a subjective mismatch between environment and a child's expectations about said environment, is a *discrepancy* (Figure 4.1, c). Thus, a discrepancy cannot, by definition, be an objective entity that exists in the world independent of a child and his/her experiences. Instead, it refers to a *situation* or *state of affairs* that may occur while a particular child is interacting with a particular environment.

4.3.1.1 Discrepancy subtypes

There are two mutually exclusive subtypes of discrepancy, *surprise* and *novelty*. Each subtype represents a different type of inconsistency between a) the environmental aspect and b) the child's knowledge and expectations about it. Where the discrepant aspect is recognisably something about which a child has prior knowledge and expectations, but that she now considers to be different, unusual, or "rule-breaking" in some way, this is a *surprise*. This sub-type is divided into *surprising events*, and *surprising non-events*,[6] in which something is different or rule-breaking because it is absent, unresponsive, or *fails* to happen as expected. Conversely, if a child considers an aspect

[3]Henceforth referred to simply as the environment.

[4]Several earlier publications on this work Alcorn et al. (2013b,a, 2014) described discrepancy as the child detecting a mismatch between his expectations and the environment, rather than in terms of subjective inconsistency. These publications sometimes also called the child's set of knowledge and expectations a "mental model" of the environment. The underlying idea of discrepancy was the same, even if the explanation was different. However, this original terminology was confusing to some reviewers. A "mismatch" may have sounded like something that can objectively exist in the world, and that a child can "miss a chance" to notice or act upon. The current explanation of discrepancy seems to better communicate the idea.

[5]E.g. appearance and properties of digital objects, character behaviours, activity goals, timing, sound effects, etc. See section 4.5.2.1 for a definition of *aspects*.

[6]These are henceforth referred to simply as surprises and non-events.

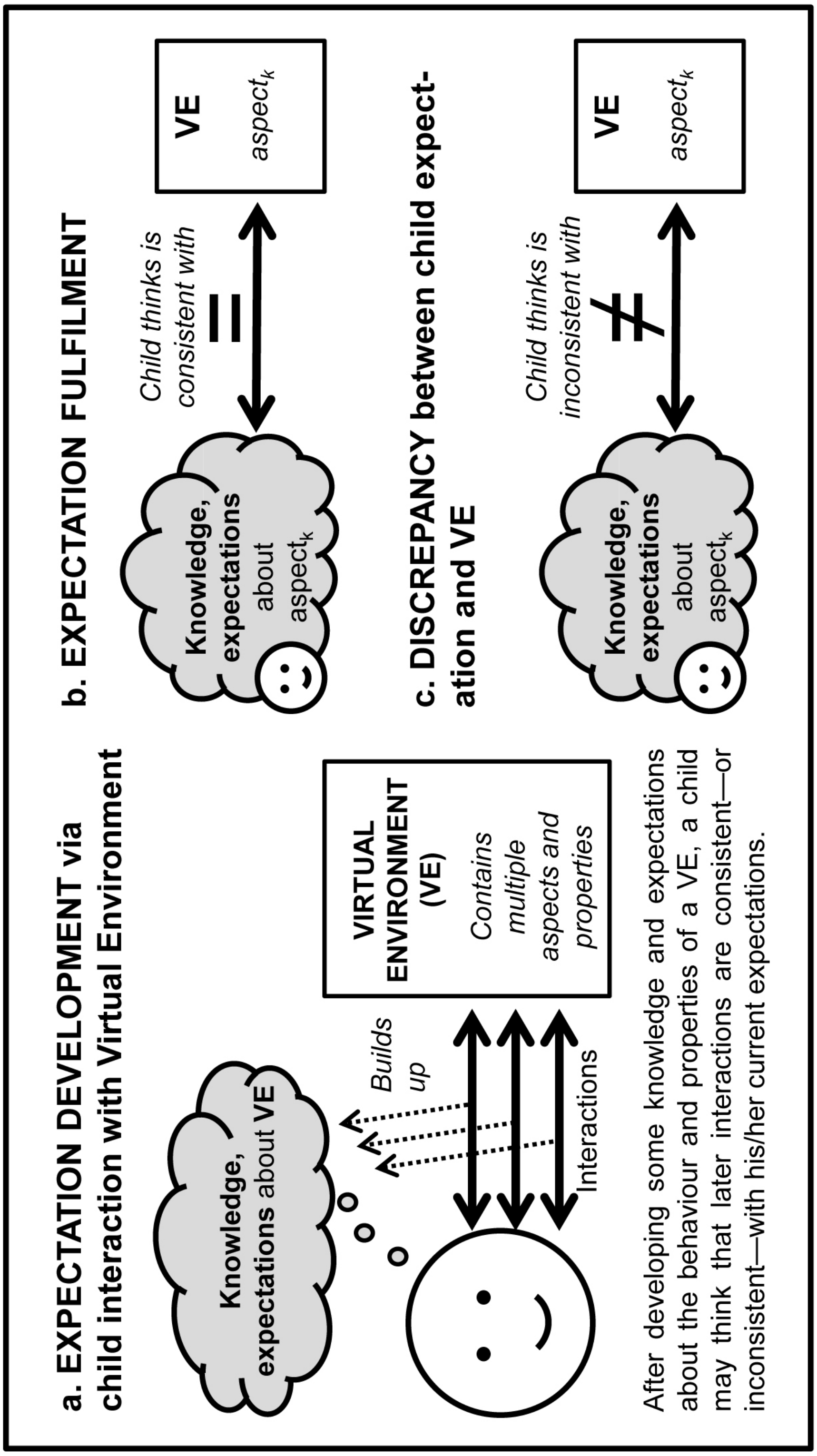


Figure 4.1: (a) Through multiple interactions, a child builds up a body of knowledge and expectations about the VE, its aspects and properties. The child's knowledge and expectations will continue to evolve and expand through later interactions. (b) When the child thinks that aspects of the VE are consistent with his/her current knowledge and expectations, this is an expectation fulfilment. (c) When the child thinks that an aspect of the VE is not consistent with (or not included within) his/her current knowledge and expectations, the situation is a discrepancy.

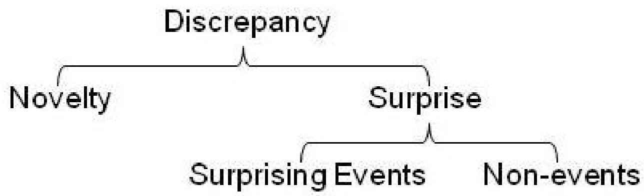


Figure 4.2: Diagram of the relation between the core category of *discrepancy* and its subtypes, in the final version of the taxonomy. The *surprise* category has not been used directly in annotation; discrepancies have instead been labelled more specifically as surprising events or non-events.

inconsistent because it is unknown and unexpected, i.e. it is absent from her current body of knowledge and expectations, then the discrepant aspect is *novel*. For simplicity, this work generally refers to three subtypes of discrepancy (surprising events, non-events, and novelty) as the category "surprise" is never annotated directly or used directly in any of the analyses. See Section 4.5.2.1 for the detailed definitions of novelty and surprise, as used in the annotation scheme.

4.3.1.2 Discrepancy as a core category

Discrepancy is the *core category* of this analysis.[7] A *core category* is an over-arching, umbrella category that can explain the main idea or theme of the analysis in a relatively condensed way (e.g. Glaser and Strauss, 1967; Strauss and Corbin, 1998; Saldaña, 2009). Charmaz (2006) sees the core category as a spine for the analytical skeleton, structuring and connecting all of the other categories. Discrepancy is a concept that was deliberately constructed as a core category to explain both of the subcategories, novelty and surprise, and put a name on their underlying similarities. Their relation is illustrated in Figure 4.2. Section 4.4.3.2 describes how the construction of discrepancy came about.

4.3.2 Taking a child-centred perspective

It is worth noting that discrepancy is defined in a child-centred manner, rather than an environment-centred one: with respect to judgements about inconsistency, the child is "always right". Some aspect (e.g. the virtual character's hello sequence) may be

[7] Sometimes also called a core code, or theoretical code.

objectively identical to those previously present in the environment, but if the child thinks it is different (i.e. inconsistent with his/her knowledge and expectations), it is a discrepancy[8]. Conversely, an aspect might be objectively novel compared to what a child has previously encountered in the environment, but if—to the child—it seems consistent with his expectations, then it is not a discrepancy, but rather an expectation fulfilment. A child's subjective perception of in/consistency is always the deciding factor in whether a situation "counts" as a discrepancy. For example, applying incorrect touch-screen actions (such as scratching or poking) led almost all children in studies 1 and 2 to believe that objects were "broken" or "not working". The environment was operating correctly, but from the children's view there was a definite inconsistency between the *expected* result of their touch action and an object's failure to respond (i.e. a *non-event*). Another example from Study 1 would be Lucy[9] commenting repeatedly and excitedly to the researcher about the "big flower" she had grown in one activity. The flower was not, in fact, larger than flowers grown in previous iterations of the activity, but Lucy clearly believed otherwise. In her view, there was an inconsistency between the flowers she had previously experienced, and this new flower that showed a "different" or "unexpected" size (a *surprising event*).

It follows from the definition of discrepancy that *discrepancy-detection* is the process through which a child decides that his/her expectations and some aspect of the environment are currently inconsistent. *Detection* is thus meant in the sense of becoming aware or ascertaining the presence of something, rather than actively trying to locate a "thing" in the environment. As already noted, discrepancies are situations, rather than external things. The definition and child-centred conception of discrepancy also means that there can be no "missed opportunities" to detect a discrepancy. Discrepancies are completely subjective, and *cannot* exist independently of a particular child. This last is a particularly important point and has been a source of past confusion, because other behavioural observations or assessments of autism *may* note the equivalent of "missed opportunities", where certain behaviours could or should have been used but were not.

[8]Some external manifestation of the child's thought is needed in order to infer a discrepancy, i.e. a reaction is also needed.

[9]Pseudonym.

4.3.3 Discrepancy-reaction pairs

An instance of discrepancy and a particular child reaction contingent on this aspect form the two halves of a *discrepancy-reaction (DR) pair*. These pairs are the main unit of analysis. The centrality of the child to defining discrepancy means that the unit of analysis *cannot* be discrepancy alone, as discrepancies are not things that exist independently in the world. They are a type of relation *between* the child and the environment.

The only way that a researcher may begin to infer what a (particular) child expects or knows about the environment—and its consistency or inconsistency—is through that child's observable behaviours (i.e. reactions) during his or her interaction with the environment. Child reactions allow limited inference about whether a child finds something discrepant, and which aspect is the "source" of the discrepancy (i.e. is perceived as inconsistent). However, observable behaviour is not in itself sufficient for analysing DR pairs. Behavioural evidence must be combined with a researcher's knowledge of the child, the context, and the child's "history" of interaction with the current environment. Understanding what the child has been exposed to in the environment (and how frequently) allows a researcher to make judgements about what expectations a child may reasonably have developed about the environment at a given point in time. When extending this concept to hands-on analysis, it is necessary to establish heuristics for recording when expectations have been developed (i.e. that an aspect is now familiar) to make sure this is done consistently across children (see Section 4.5.3.3 for details).

4.3.4 Child reactions, initiations, and responses

A child's *reaction* is his/her visible and/or audible behaviour, contingent on some aspect of the environment. Its form will vary depending on the individual child, and the aspect of the environment with which s/he may be interacting. A reaction may include any observable behaviours (or combination thereof) and have any emotional valence.

Child reactions to a discrepancy can be categorised as either *social* (directed to a partner) or *non-social*. In this work, the term *partner* has similar usage to the SCERTS model (Prizant et al., 2006), which defines partners as "people with whom a child engages in social exchanges in which there exists mutual influence among the partners" (Prizant et al., 2006, glossary, p.314). The sense of partner is broadened here to include *person-like* entities (i.e. the virtual character Andy) as well as people. The definition is also broadened from those with whom a child *does* have social exchanges with mutual influence, to those with who are *available* and have the *potential* to engage in such exchanges in a given setting—whether or not the child actually engages with them. Partners are frequently referred to as *social partners* in this thesis, as social and communicative behaviours are of the greatest interest to the work.

Social reactions that are spontaneous (i.e. unprompted) and purposively directed to a social partner (either adult researcher or VC) qualify as *initiations*. This is a deliberately broad view of initiation that seeks to recognise participants' communicative efforts, however unconventional they may be, and whether or not they are understood as intentional communication by the social partner. In the current annotation scheme, a child's behaviour can be annotated as an initiation even if the social partner does not appear to be aware of the child's effort, does not react to it, or misinterprets its meaning. In order to qualify as an initiation, a reaction must not be—as far as can be determined—contingent on the social partner's previous direct action(s) with or to the child. Child initiations may, however, be contingent on partner actions which are not socially directed to the child. Andy's ball-sorting mistake in Chapter 1 is an example of a non-child-directed action to which multiple children responded in a social way. A list of specific child behaviours that commonly form part of initiations (per the developmental psychology literature) are listed in Appendix B.

Socially directed child (re)actions contingent on a partner's communicative action are *responses*. These are only of interest in the current annotation scheme where they are part of a longer sequence of exchanges, which has begun with a child initiation about a discrepancy. Other social child actions that are *not* initiations may include responding to the social partner's communication (e.g. by doing a directed action), or imitating the social partner.

4.4 Development of the discrepancy taxonomy and annotation scheme

4.4.1 Philosophy and methodology of development

The development of the taxonomy and annotation scheme has been fundamentally qualitative. Qualitative methods are a good choice in the relative absence of pre-existing theory about a phenomenon or question, as is the case here. Such methods also accommodate inductive work, in which observations and individual instances are built up to new ideas at a higher level of abstraction. This has been the basic pattern of the analysis here: working from individual examples to categories that are related in a particular way, with a *core category* (sometimes called a theoretical code) that explains the major theme of the research and gives the main explanation of the phenomenon (discrepancy) in a relatively condensed way (see Glaser and Strauss, 1967; Strauss and Corbin, 1998; Saldaña, 2009).

The development of the DR taxonomy and annotation borrowed many strategies from existing methods, but has not followed a single method throughout. It is generically qualitative almost by default—few methods discuss or can easily accommodate secondary data, video, or mostly visual/behavioural (rather than linguistic) data. The goal of the current process was also different from that of many qualitative studies. The research questions did aim to describe and understand a phenomenon within a particular context, but there was always a longer-range goal[10] to use the initial ECHOES data as the basis of an annotation scheme that could be applied elsewhere, with newly collected data. Thus, whatever taxonomy was produced would be a more useful tool if it did not "over-fit" the exact contents of the current data by creating many, specific labels and categories. Working at a higher level of abstraction, still well-grounded in specific examples, would best meet the research goals.

The main business of the taxonomy and annotation scheme development was *coding* and *category development*. Many authors discuss these general ideas; Saldaña (2009) gives a good overview. In qualitative work, a *code* is generally a word or phrase that summarises or captures the sense of a piece of data (pieces may vary in size). Applying codes enables a researcher to begin organising, grouping, and separating small bits of data, based on shared characteristics. These groups of data are generally called *categories*. Depending on the stage of an analysis and the methods used, this grouping may be based on different criteria, from a more intuitive "sense" of similarity, up to specific definitions regarding category membership. Data may be re-coded and re-categorised multiple times within a project, as ideas develop. After some *initial (open) coding*, the current work could best be described as using *middle-order coding* (Dey, 1993). This term is discussed in more detail in section 4.4.3.1, in relation to the phases of development. Coding and categorisation cannot stand alone in qualitative research, and are invariably intertwined with a range of other analytic and reflective activities. Such was also the case in this work.

Grounded theory (GT) in its different and sometimes conflicting incarnations (e.g.

[10] After early exploration suggested there were enough examples to suggest discrepancy was a general, cross-child experience.

Glaser and Strauss, 1967; Strauss and Corbin, 1998; Charmaz, 2006) was an important influence, and provided useful tools for coding and analysis. As a method, GT is essentially about iteratively and inductively developing categories by "'flip-flop[ping]' between data and conceptualization" (Pidgeon and Henwood, 2004, p.630), linking them together, and describing and exploring their relationships. It is important to note that parts of GT were employed *as methods*, rather than to produce *a Grounded Theory* of discrepancy. Key methodological steps and strategies borrowed from grounded theory and other qualitative research sources are as follows:

- Writing *memos* throughout, to explore and document thought processes and decision-making about emerging categories and their relationships;
- *Theoretical sampling* as much as possible within the secondary data set (i.e. the corpus of ECHOES video data), in the form of choosing which children to analyse and in what order;
- *Constant comparison* of examples, codes, and categories (especially as new data was added) in order to seek new insights and investigate possible similarities, differences, negative cases;
- Creation and revision of written *concept and category definitions*, with "rules" for instance membership, as particularly discussed in Turner (1981); Martin and Turner (1986); Pidgeon and Henwood (2004);
- Creating *diagrams* as a way to explore and document category relationships (e.g. as in Dey, 1993);
- Overall, a move from specific instances in the data, to description, to analytic labels and categories with relevance beyond the individual data set.

Several of these strategies, as well as the types of coding undertaken, are discussed further in section 4.4.3.1, describing the activities undertaken in different phases of taxonomy and annotation development.

4.4.2 Iterative development: cycling between taxonomisation, annotation, and data

The individual categories, taxonomy, and annotation scheme were all iteratively developed over a long period of time, by working back and forth between draft concepts,

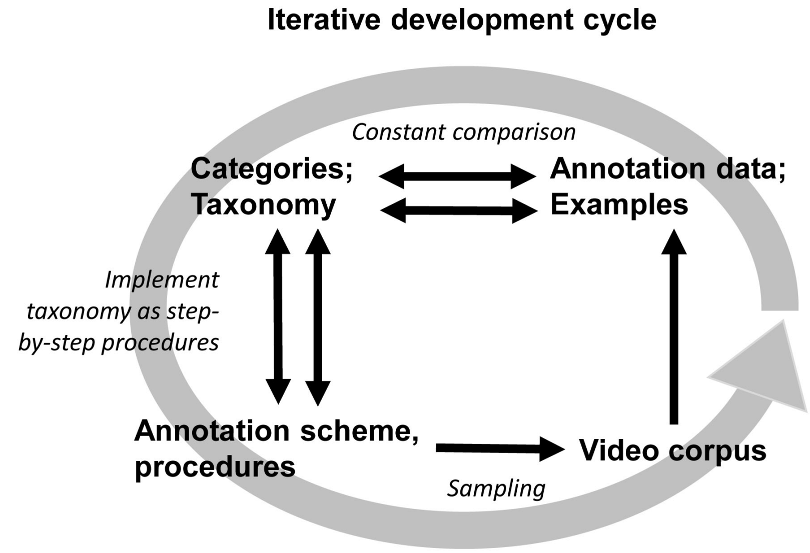


Figure 4.3: Cycle of iterative taxonomy and annotation scheme development. The cycle approximates the steps undertaken in phases 1-4.

annotation labels, and subsets of an existing data set. This process is summarised in Figure 4.3. Each cycle roughly corresponds to one phase of development, though each cycle and phase required different strategies as the work evolved. The development can be divided into five phases in total, including a phase 0 (initial ECHOES observations, not corresponding to a cycle).

The annotation scheme was a concrete implementation of the taxonomy as it existed at various phases (see Figure 4.4). Authoring annotation scheme documentation—to a level of clarity and specificity as could be followed by another researcher—was an important tool for testing the completeness of categories and labels. Consequently, the taxonomisation and annotation cannot be separated and discussed in isolation when considering different phases of development. As this work was, from the beginning, interested in the conjunction of child behaviour with things in the environment, the discussion of the discrepancy and child reaction taxonomies also cannot be neatly separated in the discussion.

A key message from Figure 4.4 is that—while the analysis is always concerned with the child, the environment, *and* their relationship—the final taxonomies capture environmental information only *indirectly*. The annotation scheme collects additional descriptive information about both environmental aspects and child behaviour, as context and evidence for the reaction and discrepancy labels[11].

[11] Environmental aspects are further analysed as part of the design work in Chapter 6.

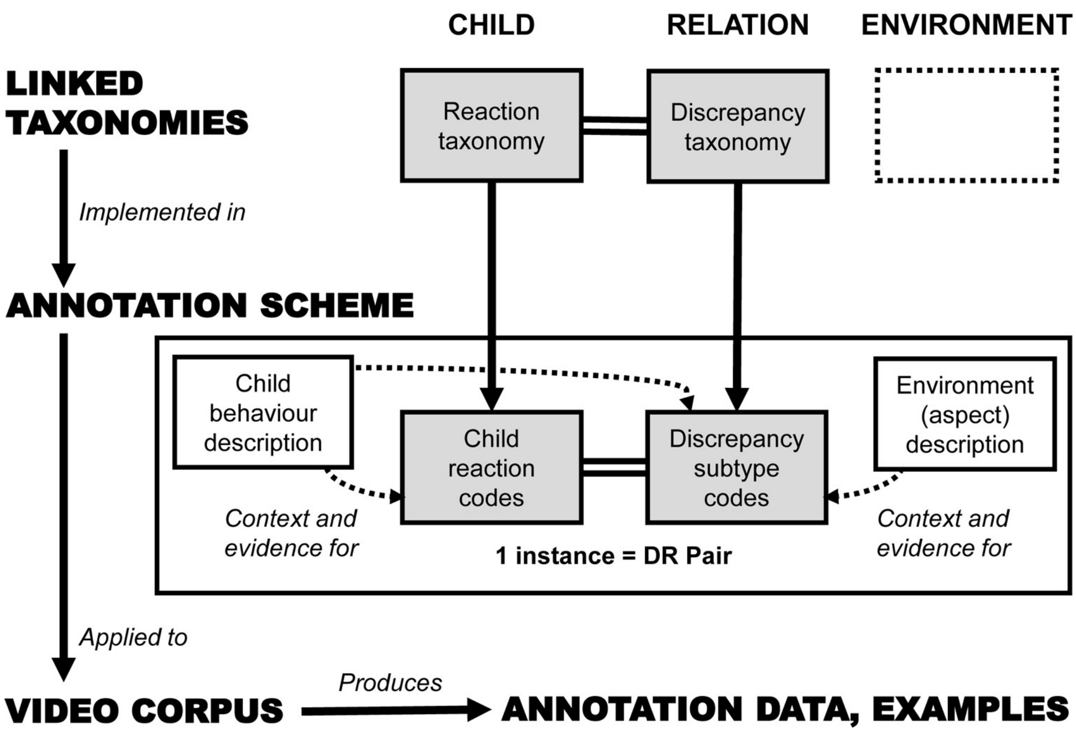


Figure 4.4: Illustration of how the three main concerns of the current analysis (child, environment, and their relationship) are represented in taxonomies, and implemented in the annotation scheme.

4.4.3 Methodology and conceptual shifts during phases of taxonomy development

4.4.3.1 Key methodological activities and outputs, by phase

The taxonomy and annotation scheme development can be divided into roughly five phases: initial ECHOES observation, exploration and initial coding, a “proof-of-concept” analysis[12], an expanded analysis with a first complete draft of the taxonomy and annotation scheme[13], and a final version of the taxonomy and annotation scheme, with core category (discrepancy)[14]. In the early phases, there was an important interplay between methodological and practical considerations. Phase 1 and phase 2 were very concerned with identifying (potentially) related instances and estimating their distribution across children and ECHOES sessions because it was not yet clear whether there was indeed “a phenomenon” to be investigated, or only a few isolated but interesting examples that would not form sufficient basis for taxonomy and annotation scheme development. The analysis in Phase 2 suggested that this critical mass of examples was indeed present in the ECHOES data, and thus the later phases worked more specifically towards generating categories and annotation procedures.

Table 4.1 outlines the analytic work undertaken in each phase, and the methodologies most applicable to that phase. It lists the key methodological outputs of each phase (e.g. provisional codes, category saturation). While some of the methods and steps mentioned are well known and flexibly applied (e.g. writing memos), others benefit from further explanation here.

Initial and middle-order coding (in phases 1 and 2) The coding process followed here was definitely not a completely bottom-up “line-by-line” type process, attaching labels to small pieces of video and very gradually building up to higher-level categories. The initial observations from ECHOES already suggested a possible—conceptual—relationship between different situations, which were already “chunks” of video that might encompass multiple interactions. Phase 1 revisited these and added to the list of examples, describing them and querying them for similarities and differences. This *initial coding* attached various labels, trying to capture any information that might be relevant for understanding the emerging phenomenon of discrepancy (or,

[12]Briefly reported as a conference poster (Alcorn et al., 2012).
[13]Presented in the formal thesis proposal for this PhD.
[14]As reflected in Chapters 5 and 8. Early versions of Chapter 5 results were published in Alcorn et al. (2013b,a) and contribute to Alcorn et al. (2014).

Phase	Description	Activities, methodologies	Sampling	Key outcomes
0	Initial observations, data immersion	• Pre-coding, noting striking examples • Write memos, discuss examples with colleagues	Return to examples noted in ECHOES analysis, across large corpus	• Formulate initial idea, descriptions of "what to investigate" in the data
1	Exploration; Initial, non-systematic coding	• Initial (open) coding, to seek "analytic leads", propose tentative codes (Saldaña, 2009) • Description of examples	Look at 3 children with phase 0 examples, presence of experiential data (i.e. "rich response" data sources, Morse 1991, as cited in Pidgeon & Henwood, 2004).	• "Analytic leads" for further analysis, tentative codes to apply in Phase 2 analysis • Practical: Phenomenon appeared repeatedly across children, appears to be part of an emerging pattern. I.e., enough data to analyse, use as basis for categorisation and annotation scheme development.
2	Proof-of-concept analysis	• Middle-order coding (Dey, 1993)	Re-analysis of initial 3 "rich" children using provisional codes and annotation procedure	• Revised set of provisional codes and their properties, draft explanation of similarity between instances in categories • First-draft annotation scheme, application rules
3	First-draft version of taxonomy, prior to core category	• Additional coding • Category refinement via constant comparison, category definitions • Diagram construction	Theoretical sampling in secondary data set: add 2 children from different school setting, with little experiential data	• First "full draft" of categories and their definitions • Provisional diagram of category relationships • Second-draft annotation scheme and application rules (substantially more complete and internally consistent than Phase 2)
4	Final taxonomy with core category, final annotation scheme	• Additional coding • Category refinement via constant comparison, category definitions • Diagram construction • Consultation of novelty and surprise literature • Construct core category	Add additional 2 participants to test for saturation (total 8 participants)	• Fully saturated categories (e.g. Glaser & Strauss, 1967) • Core category (discrepancy), explanation of its relevance, revised explanatory and taxonomic category diagrams • Final categories in broad agreement with, incorporate insights from literature • Complete annotation scheme and application rules

Table 4.1: Key methodologies and activities undertaken in each phase of taxonomy development, with information on outputs during or at the end of each phase.

indeed, if it *was* a phenomenon, not only a few examples).

The following analysis is well-described by what Dey (1993) calls *middle-order coding*. He suggests that it falls, appropriately, in the middle between a line-by-line approach, and a *holistic coding* approach that tries to identify major themes and ideas from immersion in a whole dataset. It means creating broad categories at a middle-level of description, and preliminarily grouping data into them. Dey notes that these first distinctions may be "fairly common sense categories", which was true of the very first memos and tentative labels, which reflect an everyday understanding of what it means for something to be novel or surprising (1993, p.110).

A benefit of middle-order coding is that it can proceed "up" to more abstract categories, or "down" to more detailed analysis of individual instances, generation of subcategories, and so forth. The current analysis worked both up and down during the intermediate phases (2 and 3), but ultimately only mid-level and high-level categories were used. Sub-categorisation of child behaviours and discrepant aspects was explored at some length[15], but did not ultimately appear very useful in terms of better understanding what discrepancy *is*, and looking ahead to discrepancy as a possible tool for future design. There was also, as mentioned earlier, a concern about "over-fitting" the current ECHOES data set with many detailed subcategories that might not be useful elsewhere. The goal was always to use existing data as a means to understand a new phenomenon *in as general a way as possible* and to create a research tool (annotation scheme) that could be applied to other video data of children and technologies. Middle-order coding proved very appropriate for meeting these goals.

Theoretical sampling (in phases 3 and 4) Theoretical sampling, especially in relation to GT, usually refers to the collection of additional primary data, in light of the categorises and themes that appear to be emerging from early analysis. This data is collected with the goal of checking, challenging, or elaborating an emergent theory. When using a large secondary dataset, it has been perfectly possible to apply the *idea* of theoretical sampling, strategically adding subsets of the ECHOES data to the analysis towards the same ends of refining and saturating the categories, or seeking negative cases (i.e. those that may not fit the analysis).

Phases 1 and 2 both began by sampling a small number of so-called "rich response" sources, or children with phase 0 examples and *experiential data*. Morse 1991 (as cited in Pidgeon and Henwood, 2004, p. 635) recommends that analysing a few of these

[15] Some preliminary behaviour sub-categorisation was reported in Alcorn et al. (2012).

sources (participants) can be a very productive way to begin analysis, as they are likely to yield many examples and insights. *Experiential data*, or the researcher's personal knowledge that she brings to analysis of a phenomenon (e.g. see Saldaña, 2009), was a key consideration here. Some participants were well-known to the author from the ECHOES summative evaluation, giving her substantial "insider knowledge" about, for example, how those children customarily acted and communicated with others (e.g. what behaviours they might use to initiate, interpreting their sometimes idiosyncratic language). Phase 3 extended the analysis to children from another ECHOES site, for whom this experiential data was missing. This was an important test of how well the annotation scheme supported identification of codeable material and labelling of reaction categories for (relatively) unfamiliar children.

Creation of diagrams (in phases 3 and 4) Where information is not organised linearly, playing with it spatially (as in diagram creation) can be a useful way to experiment with how concepts and categories might be related, and how best their relationships might be explained. Dey (1993) highly recommends diagrams for refining analysis and working out how to present its results coherently. "Diagrammatic displays", he says, "are not just a way of decorating our conclusions; they also provide a way of reaching them" (Dey, 1993, p.201). Phases 3 and 4 saw intensive use of diagrams for exactly these reasons, as a main concern of the late phases was establishing whether, and in what way, surprise and novelty were related (see conceptual shifts 3 and 4, in Section 4.4.3.2). This was eventually resolved by construction of a core category, discrepancy, which literally filled in the question mark at the top of the draft annotation flowcharts and taxonomic tree diagrams.

Writing category definitions (all phases) From early phases, there was a push to write down emerging categories in the form of definitions for the annotation scheme. These definitions were repeatedly revised as the categories were stretched and reformed through constant comparison with new data examples. Several qualitative researchers and sources (Turner, 1981; Martin and Turner, 1986; Pidgeon and Henwood, 2004) particularly commend definition-writing as a valuable tool for explicating heretofore implicit ideas or 'senses' about categories' membership and differentiation from one another. Repeatedly writing *and then applying* annotation scheme documentation was an important tool for testing the specificity and completeness of categories and labels. It also helped to ensure that the final annotation scheme—an important output of this project—would be fit-for-purpose.

Consultation of literature (in phase 4) Earlier phases made a limited consultation of existing literature on novelty and surprise (as an emotion), but phase 4 returned to the literature to better understand the nature of novelty and surprise (cognitive), in the hopes of illuminating how these concepts were connected, and whether a core category would be possible. Literature consulted was a mix of developmental psychology, education, philosophy, and creativity literature.

While the idea of expectation-violation had been present since some of the earliest analytic writing in phase zero, consulting the literature throughout this phase facilitated a crucial development (conceptual shift 3, see 4.4.3.2), by helping to formulate *both* novelty and surprise in terms of experience and expectation: surprise requires it, novelty requires its absence. That development facilitated another shift that included development of the core category *discrepancy*. A range of other sources were generally useful for reflecting on the taxonomy. The fact that these did *not* suggest major new lines of investigation or "missing" concepts supported the conclusion that the current set of concepts had become saturated in phase 4[16].

4.4.3.2 Conceptual shifts

The development of the linked discrepancy and child reaction taxonomies consisted of many small, cumulative adjustments. Zooming out to see the entire arc of development reveals several important conceptual shifts, that can be camouflaged when looking at phase-by-phase details. There were four major conceptual shifts during the phases of coding and taxonomy development:

1. From a "system centred" or "researcher centred" understanding of the environment, to a "child-centred" understanding;
2. From a wholly affective to a primarily cognitive understanding and labelling of children's discrepancy-contingent behaviours;
3. Description of both novelty and surprise in terms of a child's knowledge and/or expectations.

[16] Useful reflective sources included: Stanley on feelings, particularly Chapter XI (1895), Boden writing about surprise and creativity (1990; 1995; 2010), Bruner writing about surprise and creativity (e.g. 1962; 1976). On novelty, see especially Hausman (1975), and Berlyne (1960, Chapter 2).

4. A shift from focusing on the nature of events (i.e. the environment) that motivated child reactions, to focusing on the nature of the child-environment relationship.

It is worth noting that there were no major conceptual changes related to child initiation, although the more general concept of child *reaction* did evolve to become more cognitive and include non-socially-directed behaviour. The behaviours that "count" as initiation remained unchanged throughout the entire process, as these were inherited from the original ECHOES project and from the developmental psychology literature. The alterations at different stages had to do with how child initiations might be meaningfully categorised for an analysis of this type, and at what level (or how many concurrent levels) of detail. Table 4.2 summarises how the phases of development map to the numbered shifts above. The antecedents and implications of each conceptual shift are discussed below.

Shift 1: System centred to child centred view The shift to a "child-centred" understanding could alternately be described as a shift from viewing and describing the environment in relation to the researcher/designer's complete and omniscient knowledge, to describing it in relation to the child's incomplete and discovered knowledge. At the end of Phase 2, there was an explicit decision to switch the analysis from being guided by what the environment "should" or was "designed" to do, to what the *child* thought it should do, per his or her previous experience. A system-centred view of discrepancy has major shortcomings, first of all that it relies on a researcher/annotator with an intimate knowledge of an environment and how it was intended to work. The researcher's viewpoint is also very different from the child's: she has prior, external system knowledge that a child technology user does not, and is also an adult.

The system-centred view also cannot easily accommodate cases in which a child does *not* react to objective error—would this then be a "missed opportunity" to show surprise? It also would suggest the child is "wrong" if s/he seems to perceive a discrepancy where, objectively, the environment is the same. As already described in Section 4.3.2, a child-centred perspective precludes counting missed opportunities, because the child is "always right" in his or her perception of the environment. This does not mean that system errors become irrelevant. They still are instances of the environment not behaving according to its usual rules, and are thus prime candidates for a child to consider them discrepant. Phases 3 and 4 continued to work out the implications of this shift, in the taxonomy and annotation scheme. As a consequence of this shift, the

Phase	Main focus of analysis	View of environment	View of novelty and surprise	Relation between child and environment	View of child's behaviour	Unit of analysis
0-2	Description, qualities of environment	System (researcher) centred, objective and universal	Novelty defined in terms of newness, surprise loosely defined in terms of expectation violation, causing surprise (reaction)	Child is surprised by environment [emotion].	Shows surprise [emotion] contingent on environment; Interest in initiations only	Surprising event
3	Description, qualities of child and environment in conjunction	Child-centred, subjective and individual	Novelty defined in terms of newness, surprise in terms of expectation-violation	Child reacts to (subjectively) surprising event	Child reacts contingently to environment; Interest in social and non-social reactions	Surprising event- child reaction pair
4	Description, qualities of child-environment relation	Child-centred, subjective and individual	Both novelty and surprise defined in terms of child's knowledge/ expectation	Child detects subjective inconsistency between his/her expectations and environment	Child reacts contingently to environment; Interest in social and non-social reactions	Discrepancy-child reaction pair

Table 4.2: Views of the environment, child behaviour, and their relationship, by phase of development.

relevant unit of analysis also changed from an event, to an event-child reaction pair. The exact nature of this pair evolved further in later phases, to reflect the shift from analysing events to analysing environment-child relations.

Shift 2: Affective to cognitive view of child behaviours Essentially, this shift acknowledged that child reactions did not have to be a child "showing surprise" or "being surprised" in an everyday sense. The coding in Phase 2 identified many examples that seemed to be a part of the phenomenon of interest, but in which children's reactions were not primarily affective. They also asked for information, tried to solve problems, gave the character directions, and so on. Sometimes the same children showed a mix of behaviours to the very same events. Comparing these instances to the (mostly-implicit) category definitions led to expanding the sense of what constituted a child reaction, to include reactions that were more "cognitive" and may not have an obvious affective component.

Shift 3: Describing novelty and surprise in terms of expectation This was a specific and very important shift in Phase 4, catalysed by consultation of literature on novelty and surprise (discussed more in Section 4.4.3.1). While a number of authors define surprise in terms of expectation-violation, formal definitions of novelty (beyond "newness") are rarer. Hausman (1975) and Charlesworth (1969) were particularly useful for helping to re-cast novelty in terms of unexpectedness or un-predictability, and work out exactly what that meant in terms of a child's prior experiences. With this shift, novelty and surprise were described in the same terms for the first time, making it a much shorter step to developing the core concept of discrepancy, that captures their underlying similarity.

Shift 4: Focus on child-environment relation Shift 3 posed the question: if surprise is the violation of a child's expectations and novelty is the absence of expectations, is there an underlying similarity or factor that would explain both situations? The data itself strongly suggested that there would be, because children's reactions to novelty and surprise, or the way they seemed to treat these occurrences, was so remarkably similar. The answer is that in both surprise and novelty, there is a particular *relation* between the child's current knowledge and expectations, and the world. The relation is one of difference: those things do not match up. Moreover, this difference is *between* the child and environment, rather than being "located" in either one[17]. It is a *state of affairs*, rather than a thing.

This insight drove both the construction of core category *discrepancy*[18] to describe and explain this underlying idea, and also instantly switched the focus of the discrepancy taxonomy. Novelty and surprise are not *types of things* in the environment, but rather *types of child-environment relations*.

4.5 Final discrepancy-reaction annotation scheme

4.5.1 Overview of annotation scheme and procedure

As explained in relation to the annotation scheme development and illustrated in Figure 4.4, the current analysis is always interested in three things: the child, the environment, and the relation between them. The annotation scheme explicitly labels and categorises child reactions and discrepancy relations, but does *not* explicitly categorise and label states of the environment[19]. It does collect descriptive information about the environment (both the "what happened" of a discrepancy, and general context) and about specific child behaviours constituent of initiations or non-social reactions, and the apparent target partner of initiations.

Annotating videos of child-technology interactions using the DR annotation scheme is really two tasks: 1) determining whether a video segment may include a DR pair, and if it does, 2) determining the discrepancy subtype and type of child reaction in that pair. These tasks are repeated over and over, for every potentially codeable segment. There are two main strategies for locating DR pairs, outlined in Section 4.5.3.2, but these are definitely heuristics rather than algorithms. Taking a child-centred view means that *anything* has the potential for discrepancy, and an entire video sample must always be watched with a careful eye to what the child appears to be experiencing. Task 2, labelling, is better defined, and supported by detailed decision trees to determine which of the mutually exclusive labels to apply to a child's reaction and the discrepant situation (i.e. the type of discrepancy). Judgements in both annotation tasks are based on detailed, contextual, descriptive information about the environment, the

[17]With additional thanks for this insight to a timely reading of Camus, and his discussion of absurdity as being located between a person and the world.

[18]I am inclined to agree with Charmaz's (2006) social constructivist account of GT as a process in which the researcher *constructs* categories through work with the data, rather than "discovering" some pre-existing theory hidden in the data, as Glaser and Strauss would have it (e.g. 1967).

[19]The design-focused work in Chapter 6 returns to this environmental information.

child's re/actions, and how these two do (or do not) appear to be contingent. For every DR pair, the current annotation scheme records descriptions of the environment, child actions, and (where relevant) social partner actions.

The current analyses (Studies 1 and 2) used an annotation program called ELAN (Max Planck Institute for Psycholinguistics, 2012; Brugman and Russel, 2004) in order to record both descriptions and labels. ELAN allows an annotator to select a segment of video, and fill in multiple "tiers" of information attached to that segment (see Figure 4.5). The annotation scheme is implemented as an ELAN template, which is essentially a blank form with pre-labelled tiers for different information, and drop-down menus of relevant category labels. The current scheme is in no way specific to the features or organisation of ELAN, and could in the future be implemented in a completely different program, collecting the same information and applying the same labels.

Completed annotations were exported from ELAN as tab-delimited text and further analysed in a standard spreadsheet program. The export options were selected such that each segment of video with attached annotations (i.e. each DR pair) became one row of a spreadsheet, with columns indicating different types of recorded information (e.g. description of child re/action, child reaction label, target partner). Inter-rater agreement for the presence of DR pairs and application of discrepancy and reaction labels was calculated as a part of Study 1. Post-annotation work and inter-rater agreement are both reported in Chapter 5.

4.5.2 Annotation label definitions and decision trees

4.5.2.1 Discrepancy subtypes

As introduced in Section 4.3.1, there are three subtypes of discrepancy: novelty, surprising events, and [surprising] non-events. Their relations to one another are repeated in Figure 4.2. The category labels are mutually exclusive. For an individual, something that is novel (i.e. of a new kind, about which there are no expectations) cannot also be surprising, as surprise requires specific expectations. Something must be either an event or a non-event. This section gives definitions of these category labels, as used in the annotation scheme. Figure 4.4 has illustrated how the labels for discrepancy subtypes are a concrete implementation of the discrepancy taxonomy (as are those for child reactions, see Section 4.5.2.2). Most of the definitions in this section are abridged versions of the definitions included in the annotation manual. Full definitions with the additional annotation guidance appear in the annotation manual (Appendix C).

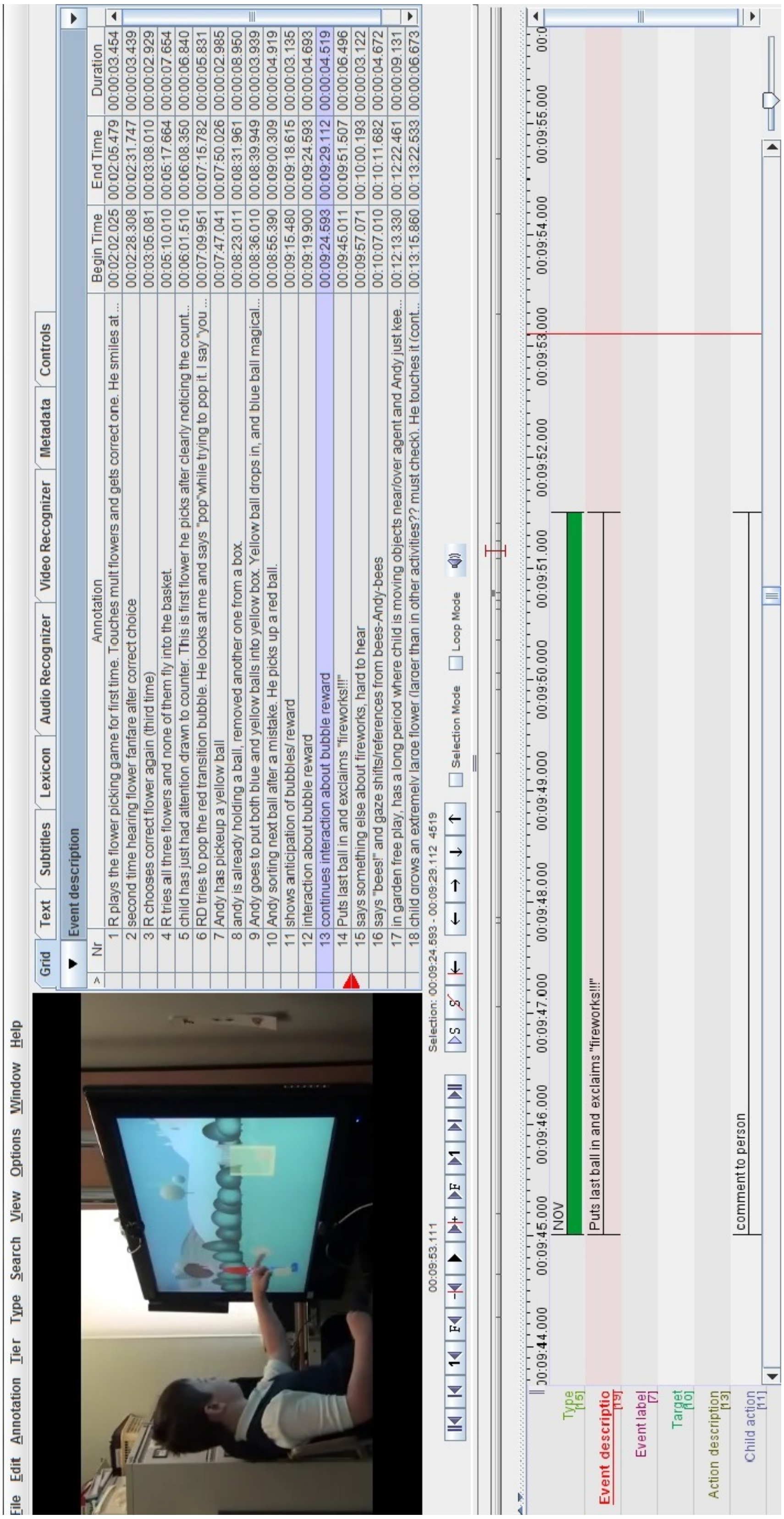


Figure 4.5: A screenshot from the ELAN annotation program, showing tiers of information reflecting phase 3 of the taxonomy/scheme development.

4.5.2.1.1 General terms used in discrepancy definitions The following general terms that are frequently re-used throughout the definitions and later analyses, to refer to various parts of the environment:

Element any piece of content within the environment, and potentially the whole environment or context itself. This might include objects and social actors in the environment. In the case of a virtual environment, an *element* can refer to the whole environment and to its various digital contents, including any virtual characters. There is no requirement for elements to be things that the user can manipulate or otherwise affect.

Event an object behaviour, character behaviour, or audio/visual effect that happens in the environment, or that a user causes to happen in the environment through his/her interaction. In some cases *event* might describe a whole sequence of actions that generally occur as a unit (e.g. Andy waves at the child and then says hello).

Non-event the actual or perceived non-occurrence of an *event*. It may also refer to the actual or perceived absence or unresponsiveness of an *element*.

Aspect an umbrella term that refers collectively to *events*, *elements*, and *non-events* in an environment. The term *aspect* is at an equivalent level of abstraction to that of *instance*, but points more clearly to the current context: individuals acting and encountering objects and events in a [virtual] environment.

Discrepancy subtype definitions also rely on the idea of *kinds* and *instances*, or universal categories and individual examples. These are particularly important for determining novelty[20].

Kind a general type, that can be considered as both a "conceptual category" and as a set of features or constituent components that hold a certain relation to one another (i.e. exhibit a certain structure). It is is universal, rather than particular. One kind is discriminable from another because the instances of that kind are typified by a certain, cohesive set of characteristics or properties that are coherently ordered and/or related to one another in a certain way.

[20]Discussion of kinds and instances draws heavily on Hausman's (1975) discussion of novelty, particularly Chapter 1, "Production and Radical Creation". The terms "kind" and "structure" are used similarly, but not identically, in this work.

Instance an individual example of a kind[21] that demonstrates enough features and relations (i.e. structure) to make it recognisable as being of that kind, rather than as belonging to another kind.

Within the ECHOES Magic Garden, **FLOWER** could be considered a kind, and any individual *flower* would be an instance of that kind. All *flowers*, regardless of size, colour, or location on-screen, would still be instances within the kind **FLOWER** because they all have the characteristics that distinguish **FLOWER** from other kinds in the environment.

4.5.2.1.2 Novelty If some aspect is the first instance of a new kind (i.e. it is novel, or has what Hausman (1975) calls "Novelty Proper"; see p.28), then the individual encountering it cannot, by definition, have predicted it based on known kinds, and cannot have expectations about it. A new (singular) instance of a known kind is *potentially* predictable, even if it was not in fact predicted. The current analyses are concerned with this unpredicted newness of kind, rather than the newness of singularity (i.e. being particularly located in time and space, per Hausman 1975, p. 20). However, it uses the simpler term "novelty" rather than "Novelty Proper". For purposes of clarity and completeness, the scheme also defines the opposite of novelty, which is *familiarity*. This concept is never annotated explicitly, but is implied in the definitions of both surprises and expectation-fulfilling events.

Novelty Novel aspects are those about which the child has not yet had an opportunity to develop expectations, such as one that is being encountered for the first time. The aspect is the first instance *k* of a new kind **K**, which has a structure sufficiently different from that of other known kinds that it could not be reasonably predicted on the basis of experience with those kinds.

***Familiarity** [22] Familiar aspects are those about which the child has already developed expectations, though these need not be complete, nor completely correct. They are instances of a known kind **K**. This category includes non-events (i.e. there are expectations about aspects' customary non-responsiveness, absence, or similar).

[21] This is roughly analogous to objects in an object-oriented programming language (such as Java) being instances of classes, and inheriting the properties of their class (i.e. their kind).

[22] Terms marked with an asterisk (*) are those defined for the sake of completeness, but not actually annotated.

For consistent application of the annotation scheme, it was necessary to create a heuristic about the (presumed) transition from novelty to familiarity. At what point could a child be expected to have formed an expectation about a new kind and its structure? The current analyses (Studies 1 and 2) employed a "rule of three exposures", though with some fine print about what the exposures should be like. See Section 4.5.3.3 for more on this rule.

4.5.2.1.3 Surprise The two sub-types of surprise both are concerned with instances of already-known kinds, those about which a child has developed knowledge and expectations. For purposes of clarity and completeness, the scheme also defines the opposite of a surprise, an *expectation-fulfilment*. This is never annotated explicitly, as it does not inform the goals of the analysis.

Surprising events are "expectation-violating events". They occur when an instance *k* of a known type **K**[23] occurs or is present in the environment, but does not appear/behave as the user expected based on his or her current knowledge of the environment and **K**. The current *k* differs from the expected *k*, but is not so different as to constitute the first instance of a new kind **K'**.

Surprising non-events (non-events) These discrepancies are defined similarly to surprising events, except that they concern violations of expectation through aspects of the environment (instances *k* of known kind **K**) unexpectedly/unpredictably being absent or failing to occur. In other cases, objects or social actors (such as a virtual character) may be present as expected, but do not perform their expected actions (based on their kind) or may not react to user actions that customarily produce a response.[24]

As with surprising events, some non-event discrepancies originate from the individual's beliefs and expectations about kinds, the environment, or the effect of his or her actions. These beliefs and expectations (etc.) may be incorrect, and *misrepresent* the actual properties of kinds or workings of the environment.

***Expectation fulfilment** The child encounters an instance *k* that demonstrates the structure and behaviours that the child expected, based on his or her previous knowledge of kind **K**. In some cases, expectations may be fulfilled by a lack of

[23]Or sufficiently similar to known instances of **K** that there could be reasonable expectations about its behaviour.

[24]In ECHOES, non-events do not mean a touch-screen failure or the entire system freezing; they are more selective issues with that affect some parts of the environment while the rest continues to function.

presence or occurrence (i.e. a non-event), such as when an element established as unresponsive does, indeed, not react to the child. Events (or feasibly non-events) may also fulfil expectations not specific to a single setting, such as when objects behave according to the physical laws of the real world and according to standard rules of causal, spatial, or object relations.

When annotating discrepancies, all judgements are being made, as far as is possible, in relation to a particular child and his/her experience within a specific environment. With respect to judgements about difference or inconsistency, the child is "always right". Some aspect may be objectively identical to those previously present in the environment, but if the child thinks it is different (i.e. inconsistent with his/her knowledge and expectations), *it is a discrepancy*. The current studies make clear that children exposed to the same material will not necessarily form the same beliefs and expectations as one another, and these may be very different than what an adult researcher understands. Many surprises may originate from an individual child's *incorrect* beliefs and expectations about kinds, the environment, or the effect of his or her actions. This may be particularly true for children who have an incomplete grasp of cause-and-effect relationships.

4.5.2.1.4 Discrepancy decision tree The relationship between the layers of discrepancy subtypes (and related categories such as expectation fulfilment) can be illustrated as a decision tree (Figure 4.6), using the language of kinds and aspects explained earlier in this section. The diagram lays out a process of comparison and decision-making as it may be followed in the annotation process, explicitly identifying the relevant criteria for determining what type of inconsistency appears to be present in a given video segment. There is no claim that this diagram represents a mental process that an *individual* may go through to detect discrepancy, and no claim that individuals have *any* step-by-step mental process to do so. The decision tree is meant as a research and annotation tool.

This diagram does not stand alone in illustrating the taxonomy. As discussed in Section 4.3.3, the unit of analysis is the discrepancy-reaction pair. Thus, in order to find a complete and analysable discrepancy-reaction pair, annotators must also consider the child's reaction (and, in some cases, its relation to the social partner's actions and prior child actions). A child reaction decision tree appears at the end of the next section.

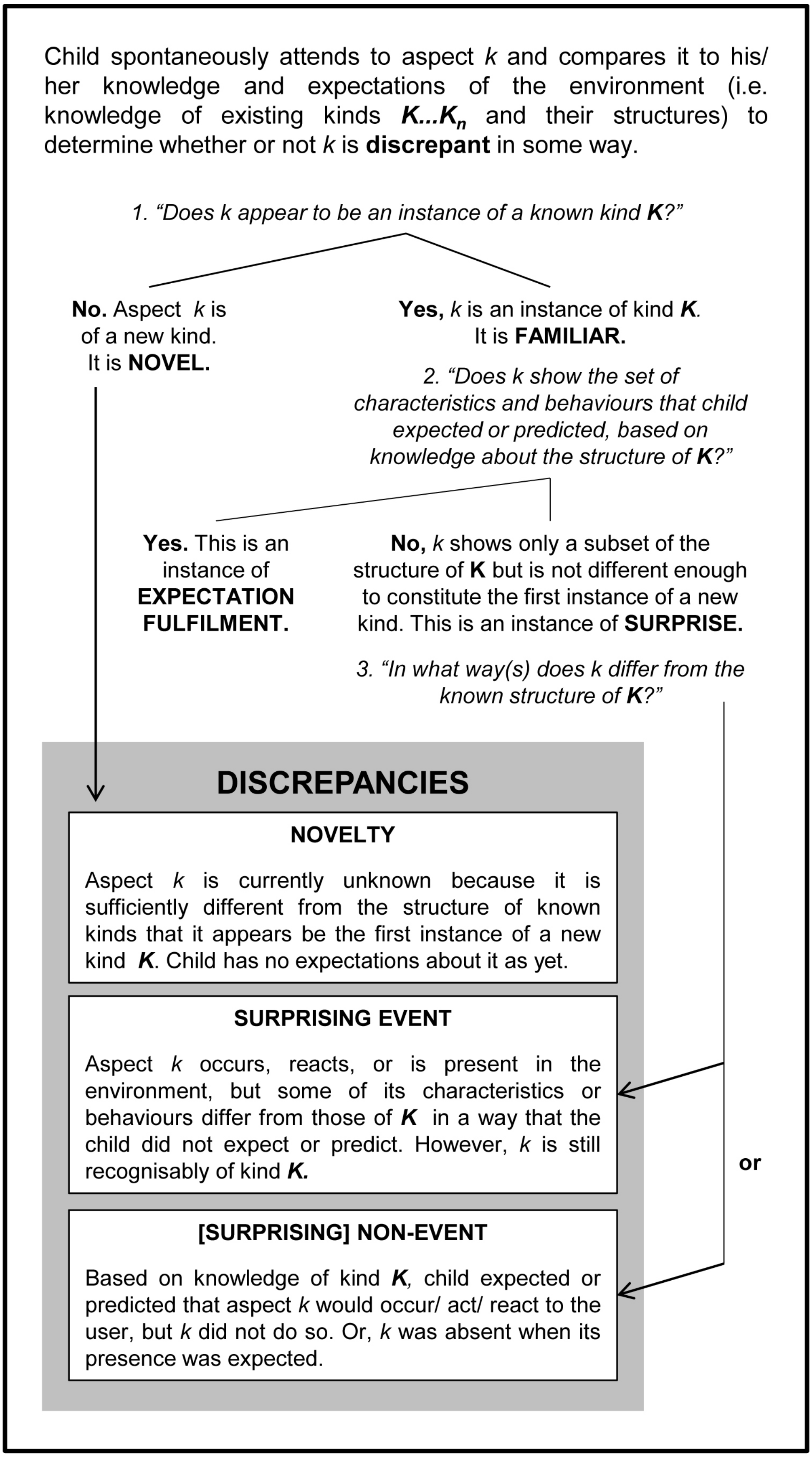


Figure 4.6: Decision tree for determining whether a child has detected a discrepancy, and what type of discrepancy it appears to be.

4.5.2.2 Child reactions

In the current analysis, a child's *reaction* is his/her visible and/or audible behaviour, contingent on some aspect of the environment. Its form will vary depending on the individual child, and the aspect of the environment with which s/he may be interacting. A reaction may include any observable behaviours (or combination thereof) and have any emotional valence. An instance of a discrepant aspect (i.e. an aspect that the child thinks is inconsistent with expectations; the source of a discrepant situation), and a particular child reaction contingent on this aspect, form the two halves of a discrepancy-reaction (DR) pair (as introduced in Section 4.3.3). The current analysis is interested in both socially-directed and non-social reactions to discrepancies.

4.5.2.2.1 Initiations The following are initiation-related labels, to be applied when the main criteria for initiation are met (see Section 4.3.4). They are not defined in terms of, nor do they label, specific constituent behaviours.

Primary initiation the first child initiation that is contingent on the occurrence of a particular discrepancy, or that has content that is explicitly about that particular discrepancy. The initiation may be made to any social partner and make take any form or combination of forms. The initiation need not occur immediately after the discrepant aspect is first encountered; it may be delayed.

Secondary initiation If the child has already made an initiation contingent on a particular discrepant aspect (i.e particular instance k), any additional initiations about that *same* discrepancy constitute secondary initiations. In situations where the child's action may be related to the presence of a partner action but not directly dependent on its form or content, the child behaviour should be considered a secondary initiation if the child is the one who is responsible for continuing an interaction that would otherwise have ended.[25]

4.5.2.2.2 Non-social reactions Not all child reactions to discrepancy take the form of initiations to a social partner. Those that are spontaneous and contingent on discrepancy but *not* socially directed are *non-social reactions*. A non-social reaction means that a child has definitely noticed an event, even if s/he has not initiated about it. These are recorded in addition to initiations because they help illuminate what children perceive as discrepant, and what they do when detecting discrepancies.

[25]See worked example at the end of this chapter for an example of a secondary initiation.

Non-social reaction When a child shows behaviour that is contingent on a discrepancy but not directed to a social partner (i.e. does not constitute an initiation), it is considered a non-social reaction. Note especially that child verbalisations should not automatically be assumed to be socially directed: they may be self-directed.

4.5.2.2.3 Child responses Finally, it is worth an explicit note that a child *reacting* to a discrepancy and *responding* to a discrepancy are not the same thing. Response has a specific meaning in the social communication literature. The annotation scheme discusses child *reactions* rather than child *responses* (except in a few special cases, as a part of multiple-turn interactions) because the discrepant aspects that appeared in ECHOES were not bids for interaction with the child. In other words, by definition the child cannot be responding because the discrepant aspects were not *directed to* the child. However, for future annotation contexts this definition may need to be amended (or another reaction definition created) to include response, should there be DR pair examples that include socially-directed discrepant aspects.

For completeness, the taxonomy defines the opposite of a child initiation to a social partner: *response to a social partner*. These instances are *only* annotated where they are part of an ongoing sequence of child-partner interaction, begun with a child initiation about a discrepancy.

***Response to a social partner; response** A child can only make a response to a partner if that partner has first made an initiation to the child. To be a response, child behaviour must be socially directed[26], purposive, and contingent on a previous social partner action that was directed to the child. The response does not need to be conventional in its form, and does not need to be noticed or understood by the social partner.

4.5.2.2.4 Child reaction decision tree The relationship between the child reaction labels and the steps for determining label application are illustrated as a decision tree (Figure 4.7). Together with the discrepancy decision tree in Figure 4.6, an annotator can identify both halves of a DR pair. As with the discrepancy decision tree presented earlier, this is an annotation tool, and there is no claim that the steps and decision in the diagram represent individuals' mental processes.

[26] As with initiation, direction may be determined on the basis of content, physical direction to/ orientation towards the partner, or extrapolation from the child's usual patterns of behaviour.

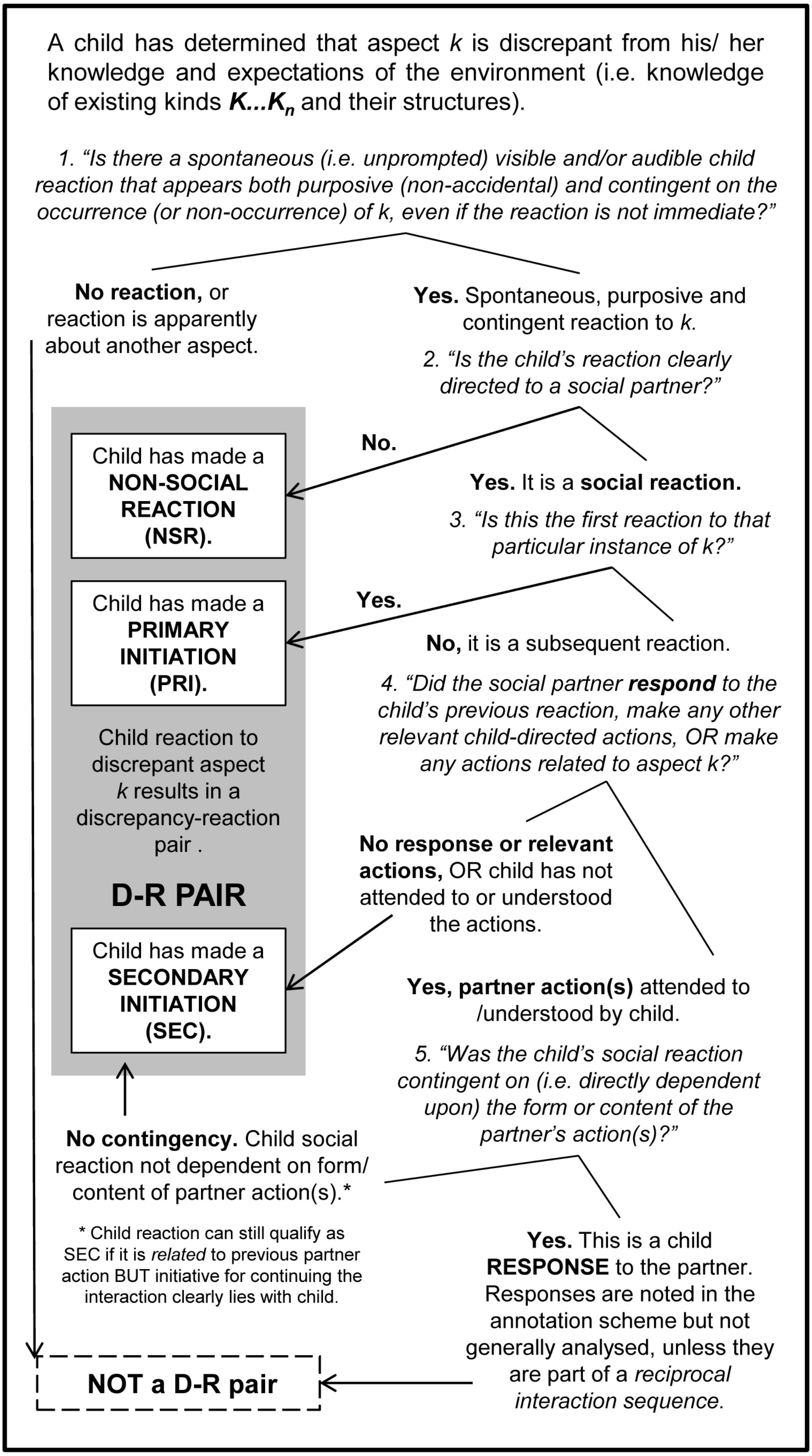


Figure 4.7: Decision for determining whether a child has reacted to a discrepancy, and what type of reaction it appears to be.

4.5.3 Annotation procedure

4.5.3.1 Annotation preparation and strategy

Prior to annotation, the researcher should collect contextual information about the data and participants. If possible, she should look at video demonstrations (or directly try out) any technologies in the video, and become familiar with them. She should also review any available research notes from the study that produced the videos. It can be particularly helpful to seek out the following kinds of information:

- General information about the study setting (e.g. a school, clinical, home).
- Information about participants, particularly descriptive information from researcher observations or familiar adults (e.g. teacher, parent). This helps give context for the child's behaviour, and may help to directly interpret it. For example, a teacher might have explained that "NAME claps when she is excited" or "he never looks at people while he is talking to them.".
- Information about communication for the child and setting: sign language, symbols, customary phrases for offering choices, phrase, instructions, etc.
- Information about what else may be happening in the room (or who else may be there) not visible to the camera.

The annotation manual advises that each participant is annotated as a unit, with all of his or her data sample (whatever that may be) annotated before going on to the next person. This makes it far easier to keep track of exposures to aspects and events (for estimating novelty), and also for developing a sense of what that child's interaction is "usually like", and thus where s/he deviates from it.

Watching a video once is not enough to annotate it well. The annotator needs to familiarise herself first, and will need to double-check for consistency at some later point. A three pass method is suggested:

Familiarisation Watch the full video segment through at normal speed, to get the gist of what happens and what the child does. If the segment for analysis does not start at the beginning of the video file, also watch the opening material.

Annotation A slow pass, frequently starting and stopping to re-watch sections and determine which labels to add. Doing this pass at 90% video playback speed can be very helpful.

Double-check A final pass to check for possibly analysable segments missed in the second pass, and also to check and make sure than similar discrepancy-reaction pairs have been coded consistently throughout the video, making changes as needed. If preferable, the double-check passes may be done for all participants once the main annotation is completed.

After annotating several participants (3-4), the first participant's annotation should be re-checked for consistency with the later participants. There will almost definitely need to be changes. The annotator may also wish to keep a running list indicating how certain types of aspects have been coded, and explaining the reasoning for coding particularly ambiguous instances in one way rather than another.

4.5.3.2 Locating DR pairs

The first step in annotating discrepancy-reaction pairs is identifying video segments in which DR pairs may be present. Taking a child-centred view of discrepancy complicates this issue, because there is no definitive list of "possible" discrepancies against which a child's reactions may be checked. This does not mean that children's perceptions of discrepancy are random or unintelligible to an observer: the data set suggests that a large number of DR pairs *do* involve something objectively wrong, missing, or different, in comparison to the system's usual operation. Many more pairs include things that are objectively novel. The structure of a DR pair suggests two strategies for finding analysable episodes. The first begins with examining the environment, and the second begins with examining child reactions.

Strategy 1: Look for objectively new aspects and objective errors or differences This strategy involves looking for things that are objectively new or different than the contents and/or behaviour of the environment up to this point (i.e. with respect to the individual child's "history" of exposure and interaction). Objective newness or difference is a useful "flag" for which data segments might merit more careful examination.

Some of the easiest discrepancy-reaction sequences to locate are those involving novelty, as defined in Section 4.5.2.1. In most cases, it will be clear whether or not a particular child has previously been exposed to a given aspect during his or her prior play, as the annotator should have kept a record about his number of exposures to different activities and parts of the environment (i.e. the *rule of three*, see Section 4.5.3.3).

After the annotator gains some familiarity with the functioning of a given environment and the content of the video dataset (i.e. during pre-coding or practice coding), it is fairly straightforward to locate things that are objectively different from the correct (or at least prevalent) behaviour of the environment. These might be actual events, such as Andy sorting a ball into a box of the wrong colour, (see 4.5.4) or deviations in the form of object absences, event non-occurrences, delays, or object/character unresponsiveness (i.e. non-events).

Strategy 2: Work backwards from child (re)actions This strategy starts with the child, identifying video segments in which s/he reacts to the environment and then working backwards in time to determine what his or her reaction was "about", and if s/he appears to find it inconsistent in some way (i.e. has detected a discrepancy). This does not mean investigating every single child action; some will clearly be responses to a social partner (see Section 4.5.2.2 for the difference between responses and reactions). Many others are play as usual, and involve the child repeating (or observing) favourite aspects/actions that are well known.

Types of child reactions that may be particularly worthy of investigation with respect to locating discrepancies are as follows (based on the ECHOES video data set and listed in no particular order):

- Child appears surprised or puzzled about something s/he has observed in the environment (any valence)
- Child appears surprised or puzzled about the result (or lack of result) from his or her own action (any valence)
- Child indicates belief that something is broken, missing, or should be happening (but is not). In some cases this is very clear, as in the case of asking where something is, or declaring that an object is "stuck"
- Child appears to find something humorous or otherwise worthy of exclamation
- Child asks for help or information about something
- Child returns to talking or asking about something that happened earlier in the session
- Child appears to unsuccessfully attempt the same action multiple times in a row

- Child abruptly shifts attention after an unsuccessful action
- Child abruptly shifts attention/action for an unclear reason
- Child appears to be watching the environment or virtual character or waiting (without touching the screen) for an unusually long time[27]
- Child verbally contradicts the virtual character or verbally/non-verbally gives him a direct instruction
- Child verbally or non-verbally attracts a partner's attention to something in the environment
- Child appears to suddenly change affect and/or level of energy (i.e. suddenly very animated and excited, suddenly seems very discouraged)
- Child suddenly approaches or withdraws from the screen (e.g. suddenly leans in very close to look at something, suddenly sits down/back and pushes chair away)

Of course, there are many types of child reactions that may be motivated by discrepancies in addition to those suggested above, and not all instances of the actions above will be related to discrepancy. They are, like objective instances of novelty, simply markers for sections of video that merit detailed examination.

Commentary on both strategies As noted in Section 4.3.1, discrepancy means that there is a certain relation between a child and the environment. This means that an annotator will almost always be working back and forth between what the environment is doing (or not!) and what the child is doing, in order to ascertain a relationship between the two. Describing the annotation piece-by-piece does not really reflect how holistic it can be. A discrepancy is a whole *situation*, and many discrepancies (in the current data sets) follow similar patterns within and across children. The annotation task in practice is largely about recognising likely situations and then checking them over in detail, with the decision trees as guidance.

[27] This varies substantially across children, and means "a long time" in relation to what the child usually does. Some children almost *never* stop touching the screen, and really struggle to inhibit their own action to watch something (such as Andy taking a turn). For such children, a four-second pause stands out in their video as very unusual. Other children may regularly stop to wait and watch things. For them, a 20-second pause may be very long.

DR pairs and the rhythm of interaction In the two datasets analysed thus far (see Chapters 5 and 8), there is not a steady rhythm of interaction between child, technology, and social partners. For example, there might be 40 seconds of near-continuous social activity when the child is having a problem or something very interesting happens, followed by a longer period in which the child is focused on the technology and makes no initiations or responses at all. The exact rhythm varied considerably between children, but did seem to ebb and flow for everyone. There often tended to be more interactions at the beginning and end of individual activities.

In both studies so far, DR pairs often clustered together—especially for children with stronger language who were better able to engage in multiple-turn interactions. Sometimes, this clustering effect took the form of a child reacting "to himself" one or more times (e.g. un-shared affect, self-narration) and *then* turning to share affect or comment to a partner one or more times. In short, if the annotator thinks there is one DR pair, s/he should check carefully before and after it, to see if there are more.

4.5.3.3 Expectation formation: the "rule of three"

As noted in Section 4.5.2.1, there is no definitive information about how much experience is required in order to develop an expectation about some aspect of the environment. In order to make sure that novelty is classified consistently within and across children, some rule is required. The suggested guideline is the "rule of three exposures", presented in an abbreviated form here:

1. A child is assumed to have formed an expectation about an aspect if s/he has been exposed to it three times. A novel discrepancy-reaction pair can only be counted if a child reacts to the aspect during one of these initial three exposures.

2. If the child continues to react to an aspect beyond those first three occurrences, the child is now assumed to have expectations about its behaviour. The aspect is no longer considered novel for the purposes of annotation.

3. Exposure is more specific than an aspect being present. It means that the child might reasonably have been able to attend to it and notice it (e.g. the researcher or child were not talking over the top of the new sound effect, the child was looking at the screen when Andy took his turn, etc.). In some cases, exposure will mean a child's direct action, such as learning about the magic cloud by trying it out.

4. If the child's behaviour indicates that s/he has already developed an expectation about an aspect in *fewer* than three exposures, then subsequent reactions to that aspect should not be counted as discrepancy-reaction pairs, even if the child has experienced the aspect less than three times in total.

To help apply this rule consistently, annotators are encouraged to track exposures within the annotation program or in annotation notes. For example, they should note how many times a character action has been seen, or if a child already shows an expectation. This counting process was officially incorporated into the annotation procedure for Study 2 (see Section 8.3.5). From both studies, this rule appears to have been a good heuristic that adequately captured most children's expectation formation and supported consistent annotation practices.

4.5.4 Worked annotation example: Initiation to researcher about a surprising event

This example is is an expanded version of an example first introduced in Chapter 1, and also a slightly modified version of one included in the annotation manual, as an illustration of applying the annotation scheme.[28] It is drawn from the ECHOES project video data re-analysed for Study 1 (Chapter 5).

Context Odell[29] is playing the "ball sorting" game,[30] in which he is supposed to help Andy sort bouncy balls (red, blue, and yellow) into the boxes of the same colours. This is his favourite ECHOES activity, and has been requested and played many times over multiple sessions. An annotator can confidently assume that Odell has formed expectations about the objects, character behaviours, and goals of this activity. With respect to locating discrepancies, there are unlikely to be any objectively novel aspects in this much-repeated game. The annotator instead should be particularly watchful for objective deviations from the activity's usual patterns (e.g. character behaviour, sounds, timing of events), and for instances where the child is reacting to ECHOES in a way that suggests he is surprised, puzzled, or has disrupted his play in some other way.

[28]Note that in some respects this is an *over*-worked example. An annotator familiar with the scheme would only need to step through the decision trees in this way for a tricky video segment, which this is not.

[29]Pseudonym of a study 1 participant; see Section 5.3.2 for further information about his characteristics.

[30]See Appendix A.

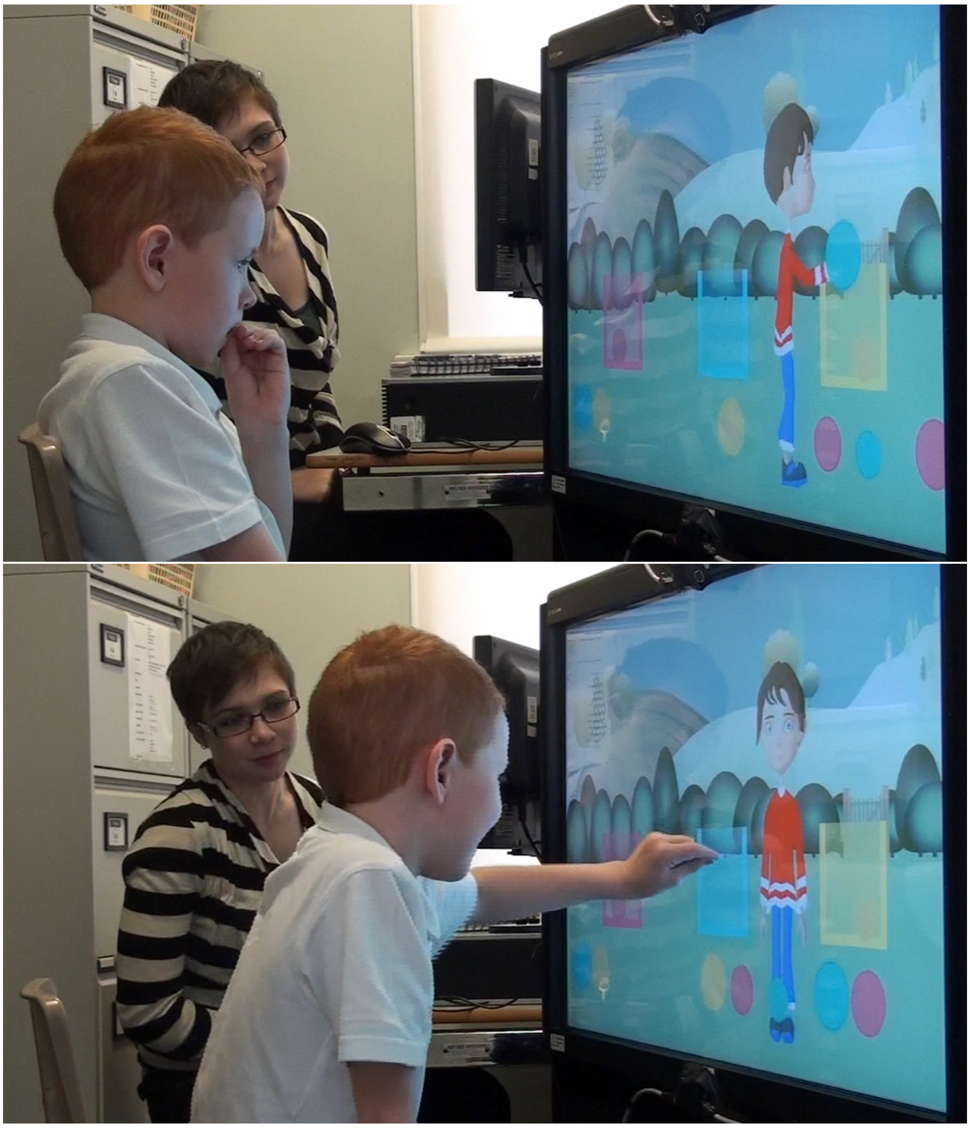


Figure 4.8: Top: Odell watches Andy make a sorting mistake. Bottom: Odell contact points to the blue box and tells Andy “Right here!”.

Description of system, child, and partner actions In this iteration of the game, after Andy and Odell each sort several balls correctly, Andy picks up a blue ball and tries to put it in the yellow box (see Figure 4.8, top). It rolls off the box top and bounces onto the ground. Andy turns to face forward (i.e. facing directly out at the child) and waits without saying anything. Odell, who is usually a calm, quiet child, has an unusually excited and animated reaction.

Researcher: *begins* "Uh-oh, Andy—"

Odell: *clearly focused on screen and not attending to researcher, he lunges forward and contact points at the blue box (i.e. touches it with extended index finger). He shouts* "Right here!" (See Figure 4.8, bottom).

Researcher: *continues* "—put it in the wrong box."

Odell: *speaking over and still ignoring the researcher, he quickly switches hands and emphatically taps the blue box twice as he shouts* "That one!" Andy does nothing. *After a brief pause Odell shifts his attention to the other unsorted balls, and begins moving a red one.*

Researcher: "Can you show Andy how to do it, and put one in that box?" *Odell does not respond, and continues his own play with a red ball at bottom of screen.*

While the researcher begins speaking about this event almost immediately, it seems fairly clear that the child is absorbed in his own interaction with Andy and is not attending to this utterance. His reaction is thus spontaneously directed to Andy (i.e. unprompted).

Applying discrepancy and reaction decision trees The definitions of child reactions ask about behaviour contingent on the environment. If behaviour does appear contingent, it is then appropriate to ask whether the reaction suggests that the child has detected a discrepancy. Here, we might go through the following thought process about this sequence, moving back and forth between considering the child, environment, and their relation.

1. Does Odell's behaviour seem contingent on some aspect of the environment? Which one?

2. Does his reaction indicate that he found that aspect subjectively inconsistent with his expectations (discrepant) in some way?

3. What type of discrepancy does this appear to be?

4. Is Odell's behaviour a spontaneous, purposive reaction, rather than one that is prompted, or a response to a social partner's communication? What type was it, and to whom was it directed?

Odell's behaviour appears clearly contingent on the environment. We have watched him watch Andy make a mistake, and immediately afterwards his affect changes. He begins excitedly talking and pointing. It seems most likely that it is Andy's failed action to which he is reacting, as has seen Andy correctly sort many balls before, using the same sequence of actions. Let us then treat the current instance as *Andy sorts ball into wrong box*, an instance of known kind or "category of things" VC SORTING ACTIONS. We can determine whether this instance counts as discrepant, and what type of discrepancy it appears to be, by following the steps in the discrepancy type decision tree (see Figure 4.6). An explanation of the relevant steps is documented in Table 4.3.

In order to determine whether a *discrepancy-reaction pair* is present, we must begin by considering the first child behaviour: the utterance "Right here!" and contact point in the second line of the description. As Odell does these simultaneously and the communicative forms reinforce one another, they are considered as a unit, rather than as two potential initiations. These actions are considered for inclusion in a discrepancy-reaction pair (and its type determined) by following the steps in the child reactions decision tree (Figure 4.7), which results in this action being recorded as a *primary initiation* (see definition in Section 4.5.2.2). Thus, there is a qualifying discrepancy-reaction pair (Surprising event- primary initiation). A step-by-step explanation of this conclusion appears in Table 4.4. Any subsequent child behaviours can then be considered afterwards individually.

Example summary The steps undertaken so far have charted how the child's reaction to an aspect of the environment, in this case an objective mistake on the part of Andy, provides behavioural evidence that he finds this event in some way inconsistent with his expectations. By using the discrepancy-type decision tree, an annotator can determine that this appears to be because the aspect of the environment was an instance of *surprise*, rather than *novelty* and was an *event* (rather than a *non-event*). The steps in the child reactions decision tree tells us that Odell made a primary initiation (PRI) to Andy. There is thus at least one discrepancy-reaction pair present about this particular

Decision Tree Step	**Step Outcome**	**Explanation for D-R example 1**
1. Does *k* appear to be an instance of a known kind **K**?	Yes, *k* is an instance of kind **K**. It is FAMILIAR (i.e. non-novel).	Child has clearly documented prior experience with this game, such that he is presumed to have formed expectations about the behaviour of the balls, and Andy's possible actions in this game, including VC SORTING ACTIONS.
2. Did *k* show the characteristics and behaviours that the child expected or predicted, based on knowledge about the structure of familiar kind **K**?	No, only a subset of the structure is present (i.e. some expectations violated). This is an instance of SURPRISE.	Andy performed a sequence of actions that was different than what was expected when he takes a turn at sorting, but still recognisably constituted an instance of a VC SORTING ACTION.
3. In what way(s) did *k* differ from the known structure of **K**?	SURPRISING EVENT. Aspect *k* occurs or is present, but some of its characteristics or behaviours differ from **K** in a way that the child did not expect or predict.	Andy does sort a ball, as expected, but puts it into wrong box (a difference).
Conclusion: Aspect was a SURPRISING EVENT. *Exit decision tree.*		

Table 4.3: Applying the discrepancy-type decision tree to *Andy sorts ball in wrong box.*

Decision Tree Step	Step Outcome	Explanation for D-R example 1
1. Is there a spontaneous, purposive reaction contingent on the occurrence (or non-occurrence) of *k?*	Yes. Spontaneous, purposive and contingent reaction to *k.*	Odell has a visible and audible reaction immediately following the surprising event. He is not attending to the researcher, and can be considered to be acting spontaneously. He purposefully uses contact pointing to indicate a specific location, and makes two verbal utterances that, based on their content, indicate that he finds something "incorrect" or "different"" about it (i.e. detects discrepancy).
2. Is the child's reaction clearly directed to a social partner?	Yes. It is a social reaction.	The utterance "Right here!" is a direction (or a correction!). Combined with contact pointing to direct attention to a location and his obvious disinterest in the researcher, Odell appears to be speaking to Andy.
3. Is this the first reaction to that particular instance of *k?*	Yes. Child has made a PRIMARY INITIATION (PRI).	The point plus utterance "Right here!" are the first reaction to the sorting mistake, though the interaction continues thereafter. Those reactions would need to be considered separately.
Conclusion: A DR pair is present (Surprising event-PRI). *Exit decision tree.*		

Table 4.4: Applying the child reaction decision tree to an example child action.

surprising event. By consulting the child reactions decision tree again about Odell's subsequent actions, we can see that he also makes another initiation to Andy (SEC), about that same mistake.

4.6 Summary

This chapter gave context for the ECHOES project video dataset mentioned in earlier chapters as having suggested the current area of research. The initial observations from ECHOES highlight the three main concerns running through this analysis: the child (or more specifically, his knowledge and expectations), the environment, and the relationship between the two. The chapter describes in detail the core category of *discrepancy*, or subjective inconsistency between a child's expectations and the environment. It explains the importance of a *child-centred perspective* to this concept, and its implications for the unit of analysis: the *discrepancy-child reaction pair* (DR pair). Child reactions to discrepancy are the main focus of this thesis.

The core category of discrepancy was constructed after a long, iterative process of analysis, building up from examples in the data toward more abstract categorisation of child behaviours and discrepant situations. The final result of this process was two linked taxonomies of discrepancy and child reactions, and a detailed annotation manual which implements the taxonomic categories and provides rules, procedures, and diagrams for identifying new instances in video data. For ease of understanding, the taxonomy/annotation development process was divided into five phases (from initial observations to final taxonomies with core category). The chapter focused on reporting key conceptual shifts that drove the development process forward, rather than giving a detailed account of each phase. This development process was informed by a range of qualitative strategies for coding and category-construction, but did not follow a single methodology from end-to-end. Key methods and analytic activities undertaken in each phase were reported, with some additional commentary on particularly useful methods.

A final part of the chapter focused explicitly on the annotation scheme, giving an abbreviated version of the reaction and discrepancy labels as applied in the scheme, and discussing the annotation procedure. Particular attention was given to the procedures for locating codeable material (i.e. potential DR pairs) within children's video data. The chapter concluded with a full worked example of a child initiating about a surprising event, using the discrepancy and reaction decision trees to illustrate how the labelling would be determined.

The annotation scheme reported in this chapter was applied to an existing data set from the ECHOES project and tested for inter-rater reliability as part of a study reported in Chapter 5. Chapter 6 returns to the concept of discrepancy in the context of its implications for the design of a new technology.

Chapter 5

Study 1: Discrepancy-reaction patterns in ECHOES video data

5.1 Introduction to this study

Observations from an evaluation of the ECHOES system suggested that there is potential for certain situations within a virtual environment (VE) to extrinsically motivate spontaneous initiation behaviour from children with autism spectrum conditions (ASC). This chapter reports a study on the phenomenon of young children with ASC (8 children aged 5-8 years) detecting and reacting to discrepancies (i.e. situations in which a child thinks a current aspect of an environment is inconsistent with his/her expectations) in a pre-existing video dataset from ECHOES. Detecting discrepancies motivated a range of spontaneous, positive initiations to both the adult researcher and a virtual character (Andy), including sharing affect, commenting, and requesting help. Nearly two thirds of reactions to discrepancies were initiations, rather than non-social reactions. These results suggest that creating objectively inconsistent situations in future VEs could provide opportunities for children to detect discrepancy (i.e. subjective inconsistency) which may, in turn, provide a useful strategy for motivating children in this group to practice initiation skills.

This chapter describes the original video data collected during the ECHOES evaluation study (re-purposed for the current study), including description of the participant characteristics, familiarisation between researchers and children, and ECHOES sessions. It also explains how the annotation scheme detailed in Chapter 4 was applied to the current data, and procedures undertaken for data auditing, second coding, and calculation of inter-rater agreement. The results report child reactions to discrepancy, focusing on initiations, and also considers the presence and causes of negative child reactions in relation to discrepancy. The chapter goes on to consider whether discrepancy might be part of a future design strategy, forming the basis for facilitating social interaction within or around technologies for autism.

5.2 Study design, research questions, goals

The current research is empirically motivated by observations of children's positive reactions to expectation-violating and unexpected occurrences during the ECHOES evaluation study (in preparation), as initially described in Chapters 1 and 4. The study re-examines the existing video data set, with its naturally-occurring examples of discrepancy, with the goals of determining:

- Whether the initially observed reactions to discrepancy were isolated occurrences, or part of a wider pattern;
- Whether such reactions were limited to specific children, or common across the participant group.

These broad goals equate to the following more specific research questions:

1. What was the quantity and frequency of children's reactions to discrepant aspects?
2. How many of these reactions constituted initiations to a social partner?
3. To which partner (human or virtual character) did children tend to initiate?

As an additional question, the current analyses and results also consider the emotional valence of child reactions to discrepancy, particularly negative reactions. This question arose through discussion and presentation of early versions of the current work, rather than being a part of the original study plan.

The focus in the current research is on exploring the basic nature of the discrepancy-reaction phenomenon in the existent data (individual young children with ASC using a VE), rather than consideration of other technologies or contexts of use, e.g. "face-to-face" social communication. It also focuses exclusively on young children with ASC. While the original ECHOES project summative evaluation also included a comparison group of typically developing (TD) children, no TD data has been included in the current analysis. As the known difficulties in social interaction associated with ASC development do not apply to TD children, TD children's data are unlikely to help us learn about the current phenomenon of interest. They also do not need specific intervention in this area. There is also no control group of children with autism, with video data in a non-ECHOES or non-technological context. The current research questions are interested in describing and understanding interactions in a particular setting (ECHOES). There is, as yet, insufficient information to guide comparative questions about discrepancy as a motivator compared to other opportunities or situations.

5.3 Methodology

5.3.1 Ethics

The ECHOES project, including associated studies and student projects, received ethical approval from the Psychology, Philosophy, and Linguistics (PPLS) ethics committee at the University of Edinburgh (2008; Reference: 160-0708). All other ECHOES partner institutions agreed to accept ethical approval from the University of Edinburgh, in lieu of application to their own local ethics committees. The ECHOES project did not require additional ethics approval from the NHS or local authorities. For the summative evaluation study referred to in this chapter, ethics procedures varied slightly at each school site, with forms customised to the names and details of participating schools. At all sites, parents gave consent for child participation and video-recording, administration of standardised measures (or access to recent scores), and specific types of data use. Children were asked to assent to their participation, commonly through verbal conversation rather than paper forms.

The original ECHOES ethics approval and parent consent paperwork extends use of the data to ECHOES researchers and associated student projects. Please see Appendix D for additional explanation of how the original ethics approval and consent for data use support use of ECHOES data in this project, with excerpts from relevant documents.

5.3.2 Participants

As the current analysis is particularly interested in children's interactions with the adult social partner (i.e. researcher) in relation to the VE, participants were drawn from the evaluation sites in which the partner's availability and style of interaction were most similar, and maximised the child's opportunity to initiate communication. Two evaluation sites were chosen as being highly comparable: they both had a social partner who was continuously, easily available to the child (in terms of physical proximity and attention), but who played a primarily reactive and encouraging role (further details in 5.3.3.2). Site 1 is a small UK primary school for children with intellectual disabilities, with an autism-specific class. Site 2 is a medium-to-large UK primary school with an autism resource base; participants spend time each day in a mainstream classroom in addition to their specialist provision. Of the other two ECHOES sites, one had a very different, more directive style of adult-child interaction where a school staff member was the main adult partner (rather than the researcher), giving frequent, specific instructions throughout the session. The equipment set-up and physical space at the final site meant that the researcher was frequently not visible and thus not easily available to the child as a responsive social partner.

Not all of the participants at evaluation sites 1 and 2 fell within the target developmental age range for ECHOES, or used verbal language. This was due to teachers requesting that all children in the same class be able to take part and try the ECHOES software. For the analysis reported in this chapter, participants were selected on the basis of being closest to the intended ECHOES project target group, i.e. with phrase-language use or better, and also with sufficient data for analysis (at least 30 minutes). This yielded 8 participants with ASC (M=7, F=1) out of the 13 total at sites 1 and 2. They had a mean developmental age of 3 years 7 months, (range 2- 5:10).

Across sites, all participants were initially recruited through informational materials distributed by their school administration on behalf of the ECHOES project. The primary recruitment criterion for the evaluation study was that each child had a previous diagnosis of an ASC by a paediatrician, child psychiatrist, or other professional. The ECHOES project did not re-confirm participants' diagnoses, but administered two standard measures in order to gain a better picture of their linguistic and social abilities: the Social Communication Questionnaire[1] (SCQ; Lord et al., 1994), completed by parents, and the British Picture Vocabulary Scale 2nd edition (BPVS; Dunn et al., 1997), a measure of receptive language ability. BPVS scores were obtained from the school's speech and language therapist (SLT) at Site 1[2], and collected by trained ECHOES re-

[1]The SCQ is based on the ADI-R (Lord et al., 1994) and often used as a screening tool for ASC. A higher score generally indicates more severe difficulties, with 15+ (out of 40) taken to indicate the presence of pervasive developmental disorder or an ASC. SCQ forms were completed by each child's parent or guardian.

[2]Scores were collected no more than three months prior to the evaluation.

Site	Child (Gender)	Age (year:months)	SCQ	BPVS: Raw	BPVS: VMA
1	Anthony (M)	7:05	23	51	5:00
	Odell (M)	7:07	25	30	3:02
	Lucy (F)	8:01	28	11	Est. 2:00[a]
	Russell (M)	7:05	23	33	3:03
2	Ethan (M)	5:05	23	44	4:03
	Hadi (M)	5:00	4 / 23[b]	30	3:02
	Kalil (M)	5:11	14	41	5:10
	Ollie (M)	4:11	16 / 26[b]	30	3:02

Table 5.1: Participant information and standardised measures (pseudonymous)

a. Child unable to complete full BPVS task. VMA estimated based on related language measures administered by the school's SLT.

b. Multiple SCQ forms completed for child. Scores from parent (L) and teacher (R)

searchers at Site 2 immediately prior to the evaluation. Table 5.1 reports participants' demographic information and test scores, including their verbal-mental ages (VMA) estimated based on BPVS scores. The disparities between VMA and chronological age suggest that all but one participant had some degree of intellectual disability in addition to ASC.

Children with similar standardised measure scores may still be very different when it comes to their actual language use and interactions. Table 5.2 gives a brief, descriptive account of the participating children's behaviours, based on both the ECHOES study video data and, for site 1, researcher observations in school.

5.3.3 ECHOES evaluation study procedure

This section briefly describes the procedures in the ECHOES project summative evaluation study, focusing on those that are relevant for understanding the video data re-used in the current study. As introduced in Chapter 4, the ECHOES VE comprises a suite of game-like learning activities[3] set in an animated "Magic Garden" and accessed through a 42" touch-screen. Andy, an autonomous and childlike VC, functions as the child's guide and playmate. The software planned Andy's behaviour both deliberatively and in response to the child's interactions (or non-interaction) with the system (Bernar-

[3] Described in Appendix A.

Site	Child	Description	Language
1	Anthony	Child appeared to enjoy social interaction; regularly and actively approached classmates and the researcher. Volunteered information about school activities. Low observed anxiety.	Child had conversational-level language and could be quite talkative. Very little echolalia. Would spontaneously attempt conversational repair.
	Odell	Appeared shy and quiet, but not withdrawn or "in own world". Better able to watch, wait than his classmates. Low observed anxiety.	Child had phrase-level language with very little echolalia. Appears to prefer non-verbal communication.
	Lucy	Short attention span, great difficulty with watching and waiting. Affectionate with adults, motivated by "helping". Low observed anxiety.	Child used words and some phrases. Some echolalia. Pronunciation was often difficult to understand.
	Russell	Seemed to passively enjoy social interaction, sometimes approached others. Low observed anxiety.	Child had phrase-level language with some echolalia. Talkative when excited, though sometimes was hard to understand.
2	Ethan	Substantial perseverative behaviour, with difficulty switching attention and tasks. Unwilling to experiment. Frequently but indirectly requested help by saying he couldn't do things.	Child had at least phrase-level language, but actual use in sessions was very inconsistent. Some echolalia.
	Hadi	Somewhat passive, very willing to wait and watch the environment.	Child used words and occasional phrases. Some echolalia.
	Kalil	Very comfortable exploring within ECHOES, trying things out. Unusually interested in Andy, compared to peers.	Child had phrase-level to conversational-level language with very little echolalia. Some repetitive questioning during sessions.
	Ollie	Appeared easily discouraged, frequently requested researcher help.	Child had at least phrase-level language with some echolalia.

Table 5.2: Descriptions of Study 1 child characteristics

dini and Porayska-Pomsta, 2013; Porayska-Pomsta et al., 2012; Foster et al., 2010). Both the learning activities and Andy's behaviour emphasised cause-and-effect relationships as well as visual and sensory elements, keeping instructions and language to a minimum.

5.3.3.1 Familiarisation, class observations, and teacher consultations

The ECHOES evaluation collected extensive information about children in their daily school context, prior to conducting any sessions with the ECHOES VE. This meant that the researchers were in school almost daily for several weeks, and able to gain substantial familiarity with the children. They observed and collected video in the classroom, on the playground, and during other school activities. They conducted structured, toy-based "pre-test" tasks with each participating child and, at site 2, also administered the BPVS. There was some discussion with teachers and supporting staff members about how best to meet individual participants' needs, but this was relatively unstructured and informal. A key piece of information from observations and teacher discussions was that the children at site 1 were *not* familiar with the widely-used simplified signing system, Makaton. This had been built into the ECHOES environment to support some of the VC's instructions and feedback, based on feedback from staff at the pilot school site. The site 1 school used a different set of signs. Going in to the ECHOES sessions, researchers knew they would need to "translate" these cues for children, using different signs or verbal language.

5.3.3.2 ECHOES sessions

After the familiarisation and pre-test data collection, each study participant individually completed several 10-20 minute sessions of ECHOES activities each week, over a six-to-eight week period. Over the course of the study, children were gradually introduced to activities with more complex goals. The main ECHOES analysis, and the current analysis, used a sample of each child's session video, rather than the entire corpus (see 5.3.4).

Sessions at each school took place outwith children's regular classrooms, in a relatively quiet space. At most school sites, this was a meeting room or office. During ECHOES sessions, children sat within reach of the large touch screen with a second adult researcher nearby at a small monitor, as in Figure 5.1. The researcher used a graphical user interface on the small monitor in order to manage the study session

Figure 5.1: At school site 1, Russell plays the “pot-stacking” game, and comments to the researcher about how each flower pot changes colour when it is added to the growing stack. The researcher is seated at the control monitor (not visible).

(e.g. choosing new activities), but s/he also provided extra support as needed. The researcher’s proximity and availability was a decision derived from an ECHOES pilot study (Alcorn et al., 2011) in which children frequently interacted with the researcher(s) and shared aspects of their experience without any prompts to do so. Indeed, these interactions about the system were often more varied and sophisticated than children’s interactions with the system and the VC. As a result, the summative evaluation deliberately encouraged communication in the “extended” learning environment of child, system, and researcher by positioning the researcher relative to the child so as to enable direct interactions between them. Interactions were not scripted. Instead, researchers were given general guidelines, the most important of which emphasised that the adult’s main role was that of responder and supporter, not as an instructor directing the child’s use of the system. A particularly common form of support was “mediation” between child and system in order to facilitate comprehension of the instructions and feedback (e.g. by “translating” these into key phrases or sign language used in the child’s classroom). At some sites, the responsive role and the system management were performed by two different researchers (see Figure 5.2).

Video data was the primary record of each child’s communicative behaviour with the system and the researcher, as the system’s automatic logging captured touch-screen

Figure 5.2: At school site 2, participant Kalil plays the “ball sorting” game. One researcher, seated slightly behind him, is available to respond and offer support as needed. A second researcher (far left) is seated at the control monitor and is responsible for advancing between activities and otherwise managing technical aspects of the session.

actions only. A second researcher recorded the child, screen, and second (supporting) researcher with a digital camcorder, positioned on a tripod approximately 1.25-2 metres away from the child, depending on the school site.

5.3.4 Materials and annotation procedure

347 minutes of ECHOES project videos (a subset of those collected for the total project) were annotated using ELAN version 4.4.0 (Max Planck Institute for Psycholinguistics, 2012; Brugman and Russel, 2004). This study used three fifteen-minute samples[4] per child drawn from chronologically early, middle, and late evaluation sessions[5][6]. This meant that each child's sample had video material drawn from at least three separate sessions with ECHOES. Samples included only those learning activities in which the VC Andy was present, so that the child had equal opportunities to initiate to both social partners. They excluded non-analysable video (i.e. system crashes[7] and

[4]Unlike in study 2 (Chapter 8) in which all of a child's session videos were used, regardless of length.

[5]This was the same as the video sampling used for the main ECHOES project analyses, and was initially chosen to enable cross-referencing between the original and new analyses. Ultimately, no cross-referencing was used.

[6]Ollie had only 32:46 minutes of qualifying video due to missed sessions.

[7]A system crash means that the entire ECHOES program or touch-screen software completely froze or shut down without warning, and the child was not able to interact with the system at all. In many

child rest breaks).

The annotation task consisted of two parts: locating, and labelling. This meant first determining each place where DR pairs occurred in the video, and then, for each pair, determining both the subtype of discrepancy represented (novelty, surprising events, or [surprising] non-events) and the type of child reaction. Free-text annotations recorded detailed information about discrepancy-reaction pairs and relevant social partner actions, including the context of each discrepant aspect, and the child's low-level reaction behaviours (e.g. pointing, commenting). Labels were applied for child reaction type (initiation, or non-social reaction). As described in Chapter 4, the annotation scheme distinguished between *primary initiations,* or the first initiation about a particular discrepant aspect, and *secondary initiations*, which were any subsequent initiations about that same discrepant aspect. The target of the initiation (adult researcher or VC) was noted, and separate fields recorded social partner responses to children's initiations, if any. The annotation procedure is discussed in detail in Chapter 4, including how qualifying DR pairs were initially identified within the video. That chapter also present a step-by-step annotation example drawn from the same dataset analysed here.

5.3.5 Analyses of annotation data

5.3.5.1 Spreadsheet analyses

Completed annotations were exported from ELAN as tab-delimited text and further analysed in the Microsoft Excel spreadsheet program. At this point, annotations were double-checked for consistency of labelling within and across children, with particular focus on checking whether child behaviours did, or did not, constitute initiations. Excel pivot tables were used to generate summary numbers for the results. *Sequences* were identified and their constituent lines of data copied to a separate "sequences" sheet for further analysis, namely checks for reciprocity of interaction. Any initiation followed by one or more secondary initiations constitutes a *sequence*. An example sequence is reported in 5.4.2.

5.3.5.2 Data auditing

The annotation scheme (Chapter 4) was developed based in the data now reported in this chapter. Once the scheme was completely finalised, it was re-applied to the data, as

cases, the excluded time involved the child waiting (sometimes for several minutes) while ECHOES or the entire computer were re-started.

preparation for the second coding process. This meant auditing all existing annotations in light of the final annotation rules, and checking for any missed episodes. This led to a number of items being added, disqualified, or re-labelled, and resulted in far greater consistency overall. Initial publications on discrepancy analyses in the ECHOES video dataset (Alcorn et al., 2013b,a) report the same pattern of results as in the current chapter, but slightly different numbers, because they use the pre-audited annotation results.

5.3.5.3 Identifying negative child reactions

As noted in section 5.2, discussion of earlier versions of this work raised questions about children's affect in relation to encountering discrepant situations, particularly the extent to which children showed negative affect. Given that the autism literature on the "need for sameness" might lead to predictions that children with ASC would experience rule-violation as upsetting or disruptive, this is an important consideration, especially with respect to whether discrepancy might play a role in future technology designs or interventions.

For the current analysis, *negative affect* is defined as including anger, frustration, annoyance, anxiety, significant confusion or appearing "at a loss", sadness, and discouragement. It was broadly defined, due in part to the difficulty of interpreting children's behaviours and often unconventional utterances. Affect of child reactions was not explicitly annotated, as this was not part of the main research questions at the time the annotation scheme and ELAN templates were developed. As noted in 5.2, this issue was identified later, in discussion of early presentation of the work. However, it is possible to derive an estimate of children's negative affect from the spreadsheet data, because of the free-text descriptions of child behaviour that often included direct comments on child affect (e.g. "child seems happy", or clues in their behaviour (e.g. "child is smiling", "child slumps back in chair").

The entire text of the data spreadsheet was searched for relevant terms that might indicate the presence of a negative affect. These included: anger/angry, unhappy/unhappiness, confused/confusion, sad/sadness, bad, grumpy, mad, can't, won't, frustrated/ frustration, unwilling, upset, anxious/anxiety. The spreadsheet was also searched for terms describing a child's posture, which (based on researcher familiarity with the videos) was a common indication of frustration or discouragement for many children. Terms included: sit down, slump, chair, lean/ lean back. These searches returned 31 child reactions that appeared either possibly or definitely negative, from their descriptions. All of these candidate reactions were cross-checked with the original videos, resulting in a total of 24 confirmed DR pairs that appeared to include negative child affect (see 5.4.4). While not as thorough as explicitly annotating affect for all annotations in the initial process, the current method seems to have been successful at estimating the presence of negative affect. Arguably, any attempt to label the affect of children with ASC in a research context (from their behaviour) will always be an *estimate* of emotional states. Many children with autism may express emotion in unconventional ways, or express little outward emotion at all. Even after consulting teachers about how their participants students express emotion, the researcher is unlikely to have deep knowledge of individual participants. Between incomplete knowledge, heterogeneous emotional expression in ASC, and the restricted medium of video data, it is inevitable that some cues for child affect will be missed. This problem is not limited to the current project, and is one for which there are no infallible solutions.

5.3.6 Second coding and inter-rater agreement

A subset of the video data used for the discrepancy analysis was selected for second coding (135 minutes, or 39% of the total). Due to the nature of the phenomenon under investigation, it was imperative to use the full data for several children, rather than sampling across children. Partial samples would mean that the annotator had only partial information about the "history" of child-system interaction, and thus what expectations a child might reasonably have developed.

The second coder had an academic background in psychology, but no particular expertise in developmental psychology or autism. The coder had very limited knowledge and experience of the ECHOES system[8], and had no personal knowledge of the participating schools or children. The second coder was trained on the annotation procedure using the annotation manual, summarised in Chapter 4 and included in full in Appendix C. The training process included hands-on practice with several segments of test-annotation on a child who was not used for the inter-rater agreement calculation. Each segment was discussed in detail between the first and second annotators, trying to 'calibrate' persistent areas of difference and identify situations in which the current guidance was unclear or insufficient. The trainee annotator then continued to the next segment. By the end of 3 test segments, a good level of agreement was being reached. This training process of iterative test-annotation and discussion (supported by a writ-

[8] I.e. received a system demo, in addition to viewing the video data.

ten manual) was the same as that used to train a pool of annotators for the original ECHOES project summative evaluation.

Inter-rater agreement (IRA) for the two annotators was calculated on 135 minutes of video data (or 39% of the total data) using a combination of Cohen's Kappa and a Kappa variant, the prevalence-adjusted bias-adjusted Kappa, or PABAK, proposed by Byrt et al. (1993). Given that each 15 minute video segment will comprise vastly more non-discrepancy-related material than discrepancy-reaction pairs, the PABAK is able to correct for large imbalances in the relative probabilities of categories (prevalence).[9]

In order to calculate PABAK for DR pairs, it was necessary to estimate a value for the number of DR pairs agreed to be *absent*, as the annotation scheme concerns itself with positive instances only. This required dividing the video into time "units" and estimating how many of those are taken up by discrepancies (as judged by one or both coders). The steps were as follows:

1. Estimate how long the average DR pair takes (see below), and set this value as one "unit" of time about which a presence/absence judgement needs to be made.

2. Calculate the total number of units in the 135 minutes of second-coded video.

3. Subtract the total number of DR pairs detected across both coders (198) from the total number of units.

4. The remaining number of units can be considered as *agreed absences*. Instances where only one coder thought a DR pair was present/absent have already been counted (in the 198).

The average DR pair was estimated to have a 7-second duration[10], and this was used as the length of the time-unit. A 7-second unit equates to 1157 total time units in the IRA video sample (each of which may or may not contain a DR pair). Subtracting 198 detected DR pairs from the total units yields an estimated 959 agreed absences (i.e. time units agreed not to contain a DR pair). Clearly, there are relatively few DR pairs in the dataset.

[9]Prevalence, as measured by PABAK, does not concern itself with "actual" differences in the underlying dataset, only imbalances in the marginal totals, on a table of observations. Values of the prevalence index range from -1 (probability of "yes" is 0) up to 1 (probability of "yes" is 100%).

[10]Study 2, which collected information on DR pair duration that allows an actual calculation, yielded an average duration figure of 5.3 seconds/pair. The similarity of the two datasets suggests that the original figure of 7 seconds was both appropriate, and slightly conservative.

The calculated PABAK for the identification of DR pairs (i.e. agreement that a particular section of video was codeable material under the current scheme) was 0.80[11], with a *prevalence index* (PI) of -0.75, meaning that it is far more probable for discrepancies to be absent than present. Agreement on labelling (discrepancy type, and child reaction) was calculated only for DR pairs which both coders agreed were present. Cohen's Kappa for labelling discrepancy types (novelty versus surprise) was 0.74; there was no need to correct for prevalence by using PABAK (PI=-0.16), as the probability of novelty and surprise were of similar order. The PABAK for judgements of children's reactions as social or non-social was 0.77 (PI=0.51), and agreement on initiation targets (human or Andy) was 0.96 (PI=0.91) (in both cases there was imbalance of prevalence of categories). For all agreement calculations, the bias index (BI) was very low ($<$ 0.05), suggesting negligible impact on PABAK values.

These Kappa/PABAK values suggest substantial agreement about *where* DR pairs are located in the video data (annotation task 1), and *what type* of discrepancies and reactions they are (annotation task 2). Employing a less conservative estimate of DR pair duration and the total "units" in the sample would have yielded an even higher level of agreement. Given the goals of the current work to analyse this phenomenon as the potential basis for a future technology-based intervention, these values also provide confidence that it is worth proceeding to more controlled experimentation about discrepancies as communicative motivators in VEs.

5.4 Child communication results

As stated in the introduction, the focus of the current study was, firstly, to determine the frequency of children's reactions to discrepancies that they identified while interacting with the ECHOES virtual environment. Of these reactions, we were interested in determining the relative proportion of non-social to social reactions (where social reactions are initiations to a social partner) and, in the case of social reactions, the target of such initiations (human partner or virtual character). As such, the analysis focused on counting the instances in various categories, rather than seeking between-child comparisons or comparing reactions to discrepancy versus reactions to other ECHOES components. The small sample size, within-child observations and lack of normal distribution necessitate non-parametric statistical tests, in this case Wilcoxon signed-ranks. All reported p values are two-tailed, with alpha values set at 0.05.

[11] Using a 7-second average duration as the basis of estimating agreed pair absences.

5.4.1 Frequency, target and type of child reactions to discrepant events

Table 5.3 gives counts of each child's DR pairs, divided by the type of reaction and the discrepancy subtype to which they were related. Overall, there were 383 child reactions to discrepancy (range=38-70, mean=47.87, SD=11.05). Of the total child reactions, 251 (64.49%) were directed to a social partner (either as primary or secondary initiations). Significantly more initiations were directed to the researcher (69.72%) than to Andy (T= 0, p=0.007). Each child's number and percentage of initiations to the researcher and Andy are listed in Table 5.4.

In this dataset, there were 311 unique discrepant aspects, a number lower than the total child reactions to discrepancy because some of those reactions were secondary initiations. By definition, secondary initiations must be *in addition to* a primary initiation about a given aspect that the child considers discrepant. Each child reacted to between 24 and 51 unique discrepant aspects (mean=38.87, SD=8.33). Of these 311 unique DR pairs, 132 (48.00%) included non-social reactions while 183 (52.00%) included primary initiations to the adult or to Andy[12]. This difference between the number of primary initiations and non-social reactions was not significant (T=9, p=0.25).

5.4.2 Sequences of interaction

Initiation sequences varied. The simplest included the child initiating once (primary), then repeating or expanding on the initiation (secondary), with or without a social partner response in the middle. Some sequences (or parts thereof) can be observed to constitute a *reciprocal interaction sequence*, or multiple-turn interaction in which child and partner exchange two or more consecutive initiation-response pairs, with at least one initiation from the child. This concept is adopted from the SCERTS framework (Prizant et al., 2005).

A number of initiations about discrepant aspects were not isolated, but constituted the first of several child initiations in a sequence. Of the 183 primary initiations, 50 (27.22%) were followed by at least one secondary initiation, constituting a *sequence* of child initiations. There were 68 secondary initiations in total (mean=8.5 per child, SD=6.34), with all children making at least two. The percentage of each child's pri-

[12] These two categories add up to 315 because there were 4 instances in which the child made a non-social reaction followed by a primary initiation about the same discrepant aspect. These 'duplicate reactions' have been removed when counting the total number of unique discrepancies in the dataset (311).

Child	Novelty						Non-event						Surprise						Total
	NSR		PRI		SEC		NSR		PRI		SEC		NSR		PRI		SEC		
Anthony		-	8	(20.5%)	9	(23.1%)	1	(2.6%)	4	(10.3%)	3	(7.7%)	2	(5.1%)	9	(23.1%)	3	(7.7%)	39
Ethan	8	(15.7%)	3	(5.9%)		-	15	(29.4%)	17	(33.3%)	7	(13.7%)	1	(2.0%)		-		-	51
Hadi	14	(29.2%)	10	(20.8%)	2	(4.2%)	11	(22.9%)	4	(8.3%)		-	6	(12.5%)	1	(2.1%)		-	48
Kalil	6	(8.6%)	7	(10.0%)	2	(2.9%)	13	(18.6%)	16	(22.9%)	10	(14.3%)	2	(2.9%)	8	(11.4%)	6	(8.6%)	70
Lucy	8	(21.1%)	12	(31.6%)	1	(2.6%)	3	(7.9%)	6	(15.8%)		-	1	(2.6%)	6	(15.8%)	1	(2.6%)	38
Ollie	3	(7.1%)	5	(11.9%)	2	(4.8%)	5	(11.9%)	19	(45.2%)	7	(16.7%)	1	(2.4%)		-		-	42
Odell	7	(17.9%)	5	(12.8%)		-	2	(5.1%)	6	(15.4%)	1	(2.6%)	11	(28.2%)	6	(15.4%)	1	(2.6%)	39
Russell	8	(14.3%)	18	(32.1%)	8	(14.3%)	1	(1.8%)	6	(10.7%)	2	(3.6%)	3	(5.4%)	7	(12.5%)	3	(5.4%)	56
Total	49	(12.8%)	56	(14.6%)	17	(4.4%)	50	(13.1%)	73	(19.1%)	28	(7.3%)	24	(6.3%)	31	(8.1%)	11	(2.9%)	383
Mean	8	(14.2%)	9	(18.2%)	4	(6.5%)	6	(12.5%)	10	(20.2%)	5	(7.3%)	3	(7.6%)	6	(10.0%)	3	(3.4%)	47.88
SD	3	(9.0%)	5	(9.8%)	4	(8.1%)	6	(10.1%)	6	(13.0%)	(4	(6.8%)	4	(9.0%)	3	(8.5%)	2	(3.5%)	11.05

Table 5.3: Child reactions to discrepancy, by reaction type and discrepancy subtype.

Child	Initiation to human	Initiation to VC
Anthony	30 (83.3%)	6 (16.7%)
Ethan	20 (74.1%)	7 (25.9%)
Hadi	9 (52.9%)	8 (47.1%)
Kalil	25 (51.0%)	24 (49.0%)
Lucy	22 (84.6%)	4 (15.4%)
Ollie	27 (81.8%)	6 (18.2%)
Odell	12 (63.2%)	7 (36.8%)
Russell	5 (55.6%)	4 (44.4%)
Total	150 (69.4%)	66 (30.6%)
Mean	19 (68.3%)	8 (31.7%)
SD	9 (14.3%)	7 (14.3%)

Table 5.4: Child initiations to the researcher and virtual character, as counts and as percentages of a child's total initiations.

mary initiations that continued to secondary (and subsequent) initiations varied considerably, from 8.33% for Lucy, up to 52.38% for Anthony.

Reciprocal interaction sequences are of particular interest because they could be considered more advanced, in a developmental sense, than single initiations or responses. An ongoing interaction requires social partners to coordinate their efforts, if they are to make contingent, relevant interactional moves. As reciprocity and scaffolding extended interactions were not a focus of the current research questions, this study did not systematically analyse child initiations to identify reciprocal interactions. This is reserved to future work (see Chapter 9). To illustrate the concept of reciprocal interactions, the following is a fairly typical example of the length and complexity of those noticed in the course of current analyses. Here, Russell has just seen an animated "buzzing bee" reward for the first time, upon completing a part of the ball sorting activity. The bees fly slowly around the screen for several seconds before disappearing.

> Russell: *Turns to researcher.* "Look, bees!" (PRI)
> Researcher: "Yeah, bees. (RESPONSE)
> Russell: *Re-orients to screen. He pulls his hand from the screen and gasps, pretending to be stung. He turns to look at the researcher, still holding his "stung" hand.* (SEC)
> Researcher: "Uh oh!" (RESPONSE)
> Russell: "Bees!" *Gaze shifts from researcher to screen (where bees are now gone) back to researcher and repeats excitedly* "Bees!" (SEC)

Researcher: *Agrees* “There were bees” (RESPONSE)

Here, the child and researcher act contingently on one another. They do stay fairly close to the initial discrepant aspect (bees), rather than having a series of exchanges that introduces new information or otherwise evolves toward a new topic. Given Russell’s developmental level and language (as estimated per standardised measures and as observed over the study), this sequence is a good demonstration of his skills, and is notable for also introducing a play or fantasy element when he pretends to be stung.

5.4.3 Rate of reaction to discrepancies

Numbers for children’s total reactions can be somewhat abstract, and are not an easy way to consider “how much” children reacted, or initiated. Considering rates of reaction per minute of video gives a better sense of this reactivity, and also takes account of the fact that one child (Ollie) had a shorter video sample than other participants. It is important to keep in mind that average rates do not reflect the actual rhythm of child-partner interaction within a session; this was often uneven, with periods of intensive interaction, and then periods of the child being absorbed in the game. Table 5.5 reports the rate at which each child reacted to discrepant aspects. The mean rate was 1.11 reactions per minute (range 0.84-1.56 reactions, SD= 0.25), or a mean of 3.3 reactions for every 3 minutes of play. Considering the mean rate of initiations only (Table 5.6), this group of children made 0.73 per minute (range 0.38-1.09 reactions, SD= 0.28), or 2.2 initiations for every 3 minutes of play. No children appear to be obvious outliers in terms of their reaction or initiation rates.

5.4.4 Emotional valence of child reactions to discrepant events

For any strategy that tries to motivate children with autism to initiate, it is important to ensure that the strategy engenders primarily positive reactions. Strategies that result in emotional dysregulation would ultimately be counter-productive. As one step in determining whether deliberately creating opportunities for discrepancy within virtual environments can be a viable motivational strategy, we examined the data to look for episodes of negative child reactions. *Episodes* encompass the interactions that appear to have led up to the negative or avoidant D-R pair(s), and subsequent interactions that either “resolve” the negativity in some way (e.g. child finally succeeds at problematic action and is now excited) or indicate a transition within the session (e.g. activity ends,

Child	Total reactions	Video (min)	Reactions per minute
Anthony	39	45	0.87
Ethan	51	45	1.13
Hadi	48	45	1.07
Kalil	70	45	1.56
Lucy	38	45	0.84
Ollie	42	32	1.32
Odell	39	45	0.87
Russell	56	45	1.24
Total	383	347 minutes	1.10 reactions/minute
Mean	47.88	43.38 minutes	1.11 reactions/minute
SD	11.05	4.60 minutes	0.25 reactions/minute

Table 5.5: Rates of child reaction to discrepancy, per minute of annotated video. Includes duplicate reactions.

Child	Total initiations	Video (min)	Initiations per minute
Anthony	36	45	0.80
Ethan	27	45	0.60
Hadi	17	45	0.38
Kalil	49	45	1.09
Lucy	26	45	0.58
Ollie	33	32	1.03
Odell	19	45	0.42
Russell	44	45	0.98
Total	251	347 minutes	0.72 initiations/minute
Mean	31.38	43.38 minutes	0.73 initiations/minute
SD	11.35	4.60 minutes	0.28 initiations/minute

Table 5.6: Rates of child initiation to discrepancy, per minute of annotated video.

Child	Number negative episodes	Negative DR pairs	Percent of child's total DR pairs
Anthony	-	-	-
Ethan	3	14	27.45%
Hadi	-	-	-
Kalil	3	5	7.14%
Lucy	-	-	-
Ollie	-	-	-
Owen	3	5	12.82%
Russell	-	-	-
Total	9	24	6.26% of all DR pairs

Table 5.7: Negative episodes and DR pairs per child, including as a percentage of each child's total DR pairs.

child or researcher changes topic of interaction). In total, there were 24 D-R pairs with negative affect appearing across 3 of the 8 participating children (Table 5.8). This represents a small number of the total D-R pairs in the dataset (6.26%). These are grouped into 9 episodes which vary considerably in length, and in the number of total D-R pairs within the episode.

Table 5.8 shows the summary of apparent causes of these negative episodes. All nine episodes involved one or more technical difficulties (typically revolving around lack of object responsiveness to touch). In the one case where the child appears to be disengaged, this seems to have resulted nonetheless from a prior technical difficulty.

5.5 Discussion

5.5.1 Frequency, target and type of child reactions to discrepant events

It is noteworthy that all 8 children in the current analysis—representing two school sites and a range of linguistic and cognitive abilities—frequently reacted to discrepancies throughout the data set. Together, they reacted to 311 unique discrepant aspects, a mean of 38.8 per child. Moreover, these reactions were often initiations to a social partner. Considering rate data, even the *least* active child made a discrepancy-related

Apparent cause of episode	Number of episodes[a]
Persistent problem with digital object manipulation, lack of activity progress	9
Virtual character interferes with child's turn twice in a row	1
Child disengaged, potentially wishes to avoid a particular activity	1
Child is being repeatedly ignored when trying to communicate with or help the virtual character	1

Table 5.8: Summary of apparent causes of negative episodes.

a. Causes are not mutually exclusive and may compound or interact with one another. Some episodes are counted more than once.

initiation every 3 minutes, with a mean of 2.2 initiations per 3 minutes across the all children. Clearly, discrepancy-related initiations were not an isolated phenomenon, but a consistent pattern across children and sessions, motivating communication over time. While it is not clear how far the current results may generalise to another participant group or context, it is very promising that discrepancy-detection appears to be "worth communicating about" for a range of children.

The particularly high number of initiations to the human social partner (70%) makes sense in light of Andy's limited capacity for response. In early sessions, most children appeared to take for granted that Andy would be able to hear and respond to them, especially after he had greeted them by name in the ECHOES opening sequence. However, there was no way for the ECHOES system to detect child speech, and no way for the child to attract Andy's attention other than by touching a task-relevant object. Initiations which attempted to secure Andy's attention were generally unsuccessful, through no fault of the children. When he failed to respond, several children appeared to overtly "test" Andy by poking his body, staring into his face, or saying his name more loudly. The responsiveness Andy *did* have, via reactive planning, did not always appear relevant because of its slow delivery. Planning and executing Andy's actions took several seconds, so that his feedback and instructions about activity-related actions were often very late in relation to the relevant child behaviour. By the time he responded, the child had moved on. For all of these reasons, participants quickly discerned that the adult partner responded more consistently, rapidly, and relevantly than did Andy. It is thus unsurprising that the majority of their commenting, shared affect, and requests for information were directed to the adult, particularly in later sessions. Overall, this pattern of interaction seems encouraging, and suggests that children were actively assessing their social partners, recognised that their own communications to Andy were generally not successful, and changed their behaviour accordingly. If Andy had possessed better capability for response (e.g. voice detection, faster and more relevant feedback) children might have continued to initiate to him, or possibly even shown a preference for interacting with him. It could be beneficial to expand researcher GUI controls to include a "rapid response" prompt for Andy, even if the response was minimal.

The current initiation results highlight numerous opportunities for adult social partners to scaffold child communication, following the child's focus of attention. This is in line with established transactional strategies such as those described in the SCERTS framework (Prizant et al., 2005). A child's non-social reaction to a discrepancy, for example, would be an ideal time for the partner to engage the child socially, stepping in and elaborating on the child's reaction, or asking a "wh-" question about the discrepancy. These partner initiations would, hopefully, help establish her or him as being positive, responsive, and available to the child. The number of secondary initiations is also promising, and points to opportunities for partners to support more advanced communication by consistently responding to children's initiations in a manner that would encourage the child and scaffold further "turns". Partner responses that invite further communication could support children in moving beyond one-off initiations and to help them practice their communicative competence in more extended, complex and potentially open-ended interactions.

5.5.2 Emotional valence of child reactions to discrepant events

The current method of estimating negative reactions (through a search of the free-text annotation and video cross-checking) seems to have provided a reasonable gauge of negative affect across the dataset: that it was relatively infrequent, and fairly mild. No negative reactions involved a meltdown, challenging behaviour or obvious distress. As noted in 5.7, only 6.26% of all D-R pairs were negative, and all of these instances of negative reactions concerned problems with the responsiveness of the technology, rather than the presence of inconsistency or change. The reactions appeared likely to "cluster" together into episodes, with a series of frustrating or discouraging interactions, and frustrated or discouraged child *re*-actions. There were relatively few episodes, with multiple negative reactions in each—rather than a pattern of isolated negative reactions. Many of these reactions appeared late in a child's ECHOES session, when fatigue, restlessness, or loss of concentration may have been playing an additional role.

Negative reactions generally included noises indicative of annoyance or agitation, demands for the researcher's help, or physically moving away from the screen and stopping interaction with it. This last appears to be an important clue about negative affect for some children, and is potentially also a self-regulation strategy. Children seem to be literally taking a breathing space between themselves and the frustrating or discouraging interaction. Kalil actually appears to directly comment on this at the end of his negative episode 3, where after failing to change the colours of some "special" balls that only Andy can use[13], he sits down in his chair (out of touching distance of the screen) and says he will just wait a minute. It is notable, however, that where children step or sit back, their withdrawal is temporary. They do not try to leave the room or ask to be finished with the session.

Interestingly, most of the negative reactions were from children at the same school site (site 2), and indeed from the same child, Ethan. Compared to other children, Ethan has a high percentage of negative DR pairs (27% of his total). He appeared to be the child with the most noticeably perseverative behaviour and most restricted interests within ECHOES. He struggled to independently switch his attention, resisted encouragement to explore the VE, and placed a priority on completing activity goals. It seems unsurprising that he should be particularly affected by problems with manipulating objects—these interfered with his goals, and he had with little or no interest in other aspects of the environment (and may have struggled to switch to these aspects, even if he had).

The concentration of negative reactions at site 2 may be coincidental, but there may be some relationship between features of that site and children showing negative affect. As the children at both sites encountered the same activity content and the same touchscreen problems[14], the difference may be in the researchers, and how they structured the sessions and interacted with the children. From the video data, there are several apparent differences in site 1 and site 2 researcher behaviour that *may* plausibly have impacted child affect, and may hold lessons for the future:

[13]This was an activity that was partially developed but not intended for use in the ECHOES evaluation. There is only one instance of this activity in the data set, and the researcher appears to have selected it by mistake.

[14]The same model of screen was used at both sites 1 and 2.

- Researchers sat further away from (and slightly behind) the children at site 2, versus immediately next to the child at site 1, and may have been perceived as less available. Compare the typical researcher positions in Figures 5.1 and 5.2.

- Site 2 researchers tended to wait for a child to request help or initiate communication before offering support. At site 1, the researchers tended to pre-emptively offer help when a child appeared to be having difficulties. This appears to be a difference in priorities: a greater value on encouraging initiation, including requests for help, versus a greater value on ensuring that children remained calm and had a positive experience.

- Site 2 researchers appeared to place greater emphasis on children completing a whole activity, and provided more encouragement and direct requests for the child to do so[15].

- Researchers at site 2 had pre-planned a series of activities to be played during a session, and generally did not ask children about their preferences. Site 1 mixed pre-planned activity sequences with child choices.

Due to the accumulation of several researcher differences, children at site 2 *might* have been more likely to *show* negative affect, due to the site 2 researchers less commonly offering pre-emptive support when a child appeared to be getting frustrated, but also more commonly encouraging activity completion—even when the activity was frustrating. If the site 1 researchers had followed a similar pattern, there might have been more negative affect. Perceived researcher availability and children's opportunity to choose their activities may also have played a role.

Overall, children in the ECHOES evaluation were remarkably determined to achieve their goals and complete activities, even after prolonged periods in which their efforts were unsuccessful (i.e. up to several consecutive minutes of unsuccessful effort). Generally, they seemed to recover quite quickly from frustration and discouragement when they had a small amount of success in an activity (e.g. successfully sorting a ball or stacking a flower pot), and/or completed it and could move on to another activity. Looking ahead to a potential future design, it is also encouraging that children's negative reactions thus far have clear causes in the environment. It would be hard to

[15]One of the days on which negative reactions occurred (Ethan and Kalil's episodes 2 and 3) was the last day of data collection at that school site. The researchers may have been understandably trying to maximise their session time with the children.

mitigate negative experiences in a future design if the designer could not understand why a child was reacting negatively.

The low levels of negative reaction to discrepancy in ECHOES suggest that overall, children found their experience to be emotionally manageable. They were frequently neutral, but sometimes extremely positive and only rarely negative. This is a very important finding for future technology designs that may try to derive "lessons" from ECHOES. It suggests that while children might react negatively sometimes, particularly to the "non-events" of objects or the touch-screen not working, they did not appear to react negatively to the repeated presence of inconsistent, unexpected, rule-violating elements. It would probably be manageable for a future design to include a similar volume of potentially-discrepant aspects as those present accidentally in ECHOES.

5.5.3 General discussion

In many respects, the discrepant aspects that appear in ECHOES share the same high-level strategy as the environmental modifications already discussed in the literature review (Chapter 2). Both constitute an alteration of a context and/or pattern with which the child is familiar, and allow the child to decide that there is an inconsistency himself, rather than having it announced. However, there are also crucial differences between the ECHOES context and existing strategies: for a start, the ECHOES system errors and novel aspects do not constitute embedded *demands* for communication in the same way as giving a child an empty juice container at snack time or a broken pencil with which to write her name. Except where they interfere with a child pursuing an activity goal, discrepant situations do not make any demands on the child, let alone embed prompts for specific communicative forms or content. The child can continue to play with ECHOES whether or not s/he notices that an aspect is novel or a rule has been broken (i.e. detects a discrepancy), and whether or not s/he protests or requests anything as a result. Instead, the current discrepancy examples could best be described as brief, motivating, and ambiguous situations embedded within the larger motivating context of play with ECHOES. Identifying discrepant aspects of the environment appears to be interesting in its own right—*so* interesting that the children in the current sample were frequently motivated to bring those aspects to the attention of social partners, comment, seek information, or simply share their amusement. In other words, the intrinsic reward of a potentially interesting or puzzling experience (i.e. detecting a discrepant aspect) may be a strong a motivator to initiate. Furthermore, the number of secondary initiations attests to the potential for these aspects to motivate *extended* interactions that involve multiple child communications and maintain contingency over several turns.

Despite the differences between existing environmental modification strategies for autism and the current ECHOES examples, the two clearly are related at a high level and may be mutually informative. This work offers further empirical evidence for the general efficacy of modifying environments to extrinsically motivate initiation, and also offers proof-of-principle that this strategy can be extended to virtual contexts, where tangible reinforcement possibilities are far more limited. Furthermore, it suggests that the resulting communicative exchanges can be broader in content and purpose than those likely to result from very concrete modifications that prompt children to request, protest, or comment about specific routines or objects.

Further research will be useful in allowing us to determine why, exactly, the opportunity to detect discrepancies appears to have so strongly motivated this participant group to make social initiations, often multiple times. It is also yet to be determined whether some types of discrepancies are generally more motivating for individual children than others, or whether particular types of discrepancy are more likely to lead to specific types of communicative behaviours. It is also an open question why these particular discrepancies were overwhelmingly experienced as positive, rather than upsetting.

5.6 Summary

This study examined an existing video data set from the ECHOES project, in which researchers had anecdotally observed that children appeared to be initiating about novel and expectation-violating aspects. The main goals of the analysis were to determine whether these observed reactions to discrepancy (i.e. subjective inconsistency) were isolated occurrences, or part of a wider pattern across the participant group and duration of the study, as well as whether reactions qualified as initiations of communication, and to whom these were directed. Additional analyses explored children's affect in relation to discrepancy. A subset of eight children with autism (M=7, F=1, mean developmental age 3 years 7 months) were selected from the participants in the ECHOES summative evaluation study. A 45-minute sample of each child's ECHOES session video was annotated according to the annotation scheme described in Chapter 4, which records and classifies children's reactions to discrepant aspects (*DR pairs*). 38% of the video was second-coded by an annotator uninvolved in the ECHOES project and without personal knowledge of the participating children. A prevalence-adjusted Kappa indicated substantial agreement about the presence of DR pairs in the video, and about the labelling of discrepancy subtypes and child reactions.

The annotation results indicate that initially observed child reactions to discrepancy were not isolated examples, but part of a pattern. Indeed, they occurred frequently and repeatedly across all participants, and across multiple play sessions with ECHOES. A large proportion of child reactions (64%) were initiations to a social partner, most commonly the researcher rather than the VC, Andy. Children sometimes made multiple initiations about the same discrepant aspect, sometimes engaging in multiple-turn reciprocal interactions with the researcher. For the very heterogeneous children in this study, all found some type of discrepancy to be "worth communicating about". This may be more generally true of a broader population of young children with ASC, and is definitely worth further investigation.

Relatively few reactions to discrepancy appeared to indicate negative affect. Overall, children appeared to find their experience neutral or positive. Where negative reactions were present, these tended to cluster together into short "episodes" of frustrating or discouraging interaction, often related to persistent difficulty with the touch-screen interface. None appeared related to the presence of novel, inconsistent, or rule-breaking material. More notable than these few negative reactions was children's high level of persistence in continuing to play with ECHOES, in the face of often-pervasive problems with the touch-screen.

The phenomenon of discrepancies motivating spontaneous child initiations does not seem to have been studied elsewhere for this population, and little appears to be known about it. Nevertheless, the results of this study suggest that it may represent a promising approach to motivating young children with ASC to initiate communication, at least within the specific context of a virtual environment. However, substantially more work is required to translate these early results into part of a design strategy that attempts to leverage discrepancy.

Chapter 6

Developing design principles to support transfer of discrepancy-contingent communication patterns

6.1 Introduction

This chapter begins the transition from the investigation of "naturally occurring" discrepancy in an existing context, to investigating deliberately-engineered discrepancy in a new context. This involves a combination of analysis, reflection, and hypothesis construction. It formulates principles to guide transfer of ECHOES' insights about discrepancy to new contexts and will culminate in the implementation and evaluation of a new design (following chapters).

The results of Study 1 suggest that children with a range of characteristics all found something motivating and "worth communicating about" when they detected discrepancies in a virtual environment. Despite occasional frustrations, they appeared to find their overall experience to be emotionally manageable and non-threatening. These exciting and completely fortuitous results raise the possibility that discrepancy-detection could be motivating for other young children with autism, and might be leveraged in other educational, recreational, or support technologies. Discrepancy-detection might have great potential to motivate practice of initiation and other social communication skills. It would be very desirable to try to re-create an ECHOES-like pattern of discrepancy-detection, in order to investigate these possibilities.

More analysis and reflection regarding discrepancy in ECHOES is required before progressing toward any new designs. This chapter aims to step through the progression from current data to future design. It reports three pieces of work, building on each other sequentially:

1. **Further analysis** of Study 1 (Chapter 5) annotation data, for the purposes of producing additional descriptive information about discrepant aspects and child-environment interactions.

2. **Reflection and synthesis** of Study 1 information, presenting *working hypotheses* (Cronbach, 1975), or local and provisional explanations, about what appears to have "gone right" in ECHOES with respect to children's positive reactions to discrepancy. This section tries to connect Study 1 findings to existing information on autism, and frames the overall goals and remit of the design recommendations.

3. **Implications and recommendations** for future designs that aim to leverage the discrepancy phenomenon to motivate communication in an emotionally manageable way. This section provides a set of higher-level principles to guide the use of discrepancy in a future design[1].

The end goal is to further understand and document discrepancy in the ECHOES context (a *sending context)*, and create design principles and recommendations that aim to support *transfer* of the discrepancy-reaction phenomenon to new autism and technology designs (i.e. application of insights in *receiving contexts*, see Lincoln and Guba, 1985; Shenton, 2004)[2]. Chapter 7 reports on the development of a small suite of new games called *Andy's Garden*, which use the principles to shape their design.

6.2 Further analysis and description of Study 1 data

Study 1 was interested in the nature of the discrepancy phenomenon at a high level, rather than in *what*, exactly, children found to be novel or surprising. Using ECHOES as input into a future design demands an account of the specific aspects of the virtual environment that were detected as novel or surprising, and also more thought about

[1] A version of the higher-level principles, with less information about their derivation, was initially reported in (Alcorn et al., 2014).

[2] The concept of transferability has been further discussed in Chapter 3.

between-child differences in interaction. This section reports three pieces of further analysis:

1. Labelling and categorising discrepant aspects in ECHOES, and looking for any emerging patterns that might inform future design

2. Interactivity of discrepancies, or, the extent to which children were detecting discrepancies through direct action versus through more passive means (observing, listening, etc.)[3]

3. Differences between the interactions observed in different ECHOES activities.

Integrated within these analyses are further consideration of between-child differences, both in terms of their quantitative results and researcher observations of their behaviour with ECHOES. The goal is to identify what things children (or at least some subset of children) appear *likely* to notice and perceive as discrepant. In other words, what were good opportunities for discrepancy-detection[4]? The priority is to generate information that can be used as concrete input for future designs.

6.2.1 Labelling and categorising discrepant aspects in ECHOES

Chapter 4 explained that the discrepancy taxonomy categorises information about types of *relationship* between an aspect of the environment, and the child's knowledge or expectations about that aspect. The types of relation are the discrepancy subcategories (novelty, surprise, and non-events). The annotation process collected descriptive information about the state of the environment for each DR pair-containing video segment, as evidence for the discrepancy labelling decisions (see Figure 4.4). Environmental information was not coded or categorised to answer research questions in Study 1, but is a valuable resource for future design, as it can answer the question of which specific ECHOES aspects were involved in the DR pairs. In other words, what did children actually detect as discrepant? This means not only exact objects, effects and behaviours, but *properties* that may be shared between aspects, such as VC involvement or sensory modality. Considering what individual children detected, alone or in comparison to peers, is another valuable design input. This information can then be used as the basis for extrapolating toward new discrepant situations in a new setting.

[3]The first two pieces of analysis (classification and interactivity) were undertaken concurrently, and thus refer to one another to some extent.

[4]The idea of designers creating *discrepancy detection opportunities* is developed in Section 6.3.2.

6.2.1.1 Aspect labelling

Analysing the annotations of descriptive, environmental information from Study 1 used a different type of coding than that used to construct the discrepancy-reaction taxonomies and annotation scheme. There, *middle-order coding* (Dey, 1993) began by grouping together chunks of data into provisional categories, rather than going "bit by bit" attaching descriptive labels, and later grouping instances into categories. Here, with little sense of what categories might ultimately emerge, it made more sense to take a completely bottom-up approach. Various qualitative methods use some version of this approach, though the focus of the labels attached may vary.

Each datum (the free-text description of the environment for a DR pair annotation) was given a single descriptive label identifying the aspect that seemed to be primarily responsible for the child's detection of discrepancy (i.e. what the discrepancy was "about"). As per Chapter 4, *aspect* is an umbrella term that refers collectively to *events*, *elements*, and *non-events* in the environment. This can include patterns, temporal information, and sequences of events that occur (or should occur) as units. Table 6.1 gives several examples of free-text descriptions[5] and labels.

The complete list of labels was aggregated across children, and labels referring to the same aspect were standardised. For example, the initial labels "Andy sorts ball in wrong box", "Andy puts red ball in yellow box" and "Andy ball sorting mistake" would all have become "Andy sorted a ball into the wrong box" (Aspect 11.2, see Section 6.2.1.3 and Appendix E). Each label then represented one *type* of aspect, which might then have multiple instances across the dataset. At this early stage, if there was any doubt about whether or not particular labels should be merged, they were not.

The final list of aspect labels represents 112 types of ECHOES aspects which could be considered to have "caused" a discrepancy-reaction pair. Of these, 48 are unique in the dataset (i.e. only one instance each), but most have several instances across different sessions and children. The number of labels is high, given that there were only 315 unique discrepancies in the dataset[6]. The number also is high given the deliberately limited set of objects, character behaviours, and activities included in ECHOES. It underscores the wide variability in the participating children's expectations and interests, and the subjectivity at the heart of discrepancy as a phenomenon.

[5]The free-text descriptions reported in 6.1 combine text from two separate tiers of the original annotation in ELAN (described in Chapter 4). The text has been lightly edited for clarity out of the original context.

[6]With a higher number of total DR pairs, due to the presence of secondary initiations about some aspects.

Subtype	[Activity] Free-text annotation data describing environment and child actions	Initial descriptive label	Final (standardised) aspect label
Novelty	**[Pot-stacking]** Child un-stacks and re-stacks pot that Andy put in place. Colours of pots change as they are unstacked and re-stacked. Child points at screen, turns to researcher, and comments "colour changed!"	Child discovers new object action or property--pots change colour	AG6.9. Child discovered that flowerpots change colour when stacked
	[Cloud raining or Flower growing] Child holds down cloud, gets thunder sound, and withdraws hand at once. Turns to look at researcher and shouts "wow!" then goes back to play.	Novel sound effect--THUNDER	AG8.5. Novel sound effect: thunder when cloud shaken repeatedly
Non-event	**[Ball throwing or Ball sorting]** Child told it is Andy's turn (by researcher). He waits from 6:18 to 6:23, then does a gaze shift between Andy and red balls (near Andy's feet). Nothing happens, so Child takes another turn.	Researcher says Andy will take turn or do a particular action, but he does nothing	AG16.5. Researcher said Andy would take his turn or do a certain action, but he did not
Surprise	**[Free play with Andy]** Child returns to “giant” blue ball, after playing with some other red balls. He lifts ball and "shows" it to Andy. He says "Andy, Andy, look it's big!" while looking at Andy’s face. A few seconds later Andy says "look, ball changed colour!"	Perceived extra-large ball	AG17.5. Perceived unusually large object: ball
	[Turning flowers into balls] Child makes a ball out of a flower, immediately throws it through the cloud. It changes colour and bounces (falls?) off-screen. He says something (unclear, Andy talking at same time), quickly gazes to researcher, and then looks back to screen. Researcher acts astonished and says "Where'd it go?"	Balls bounce/roll out of screen side (do not return)	AG18.3. Balls bounced/rolled out of side of screen and did not return

Table 6.1: Examples of free-text annotation descriptions from Study 1 (Chapter 5), their initial descriptive labels, and final (standardised) aspect labels. Numbers in the rightmost column are in the format Aspect Group (dot) Aspect number.

Before discussing how these labels were grouped into higher-level categories, it is important to be clear that the labels are meant as tools for summarising environment data. In most cases, labels were applied with reference to the researcher's understanding of the situation, and often included information that was not available to the child. The labels do not claim to represent the child's understanding or terminology, though in some cases they *may* do so, especially where the researcher's understanding of a situation was based on children's explanation of their own perception or understanding (e.g. stating that something is missing, or different). The labelling process requires judgements about which features are "driving" the child's reaction in each specific instance, aided by the child's specific behaviours. The aim of the final set of labels is to provide a starting point for designing future interactions.

6.2.1.2 Categorisation

Grouping aspect labels into categories (or *aspect groups*) on the basis of shared features made it easier to look for patterns and insights, especially as most labels had only a few instances each. Unlike the categories in the discrepancy-reaction taxonomies, these categories are completely specific to this dataset, and descriptive rather than analytic or conceptual.

These *aspect groups* (AG) were created through a bottom-up process of successively grouping individual aspect labels that appeared similar to one another in terms of the key feature(s), relation(s), or interaction(s) that the child appeared to find discrepant. Some of the resulting groups are straightforward, such as grouping all instances of the child reacting because s/he observed a [novel] visual reward in the ball-sorting game—regardless of which of the three rewards it was.[7] This became AG4, *Child action resulted in ball-sorting activity reward.* Others groups are more heterogeneous, such as *Andy violated a pattern/rule regarding his own usual behaviour* (AG12), which encompasses a variety of surprising events and non-events that involve Andy, but are not related to his carrying out a specific activity. Examples of this group include Andy inappropriately exiting the screen mid-activity, being stuck with his eyes closed, or not responding to the child's tickling action. Deciding to group aspects by similarity was not, however, sufficient to produce a mutually exclusive list of aspect groups that might serve to inform design. Thus, several heuristics were introduced to guide which features could be combined in the same AG and which could not. They

[7] See Appendix A for a description of this activity.

attempt to respect high-level patterns of analysis elsewhere in the same dataset, and maximise the information that may be useful for design. These are as follows:

1. **Discrepancy types:** The group cannot not mix novel and surprising aspects, as these are distinct in the discrepancy taxonomy (one requires expectation, one requires lack of expectation). Surprise categories may include both events and non-events.

2. **Primary actor:** As far as possible, groups should represent either discrepant aspects involving the child's actions or involving Andy's actions. Instances that involve both actors should be grouped depending on which part of the interaction appears to be the discrepant one. For example, when Andy tries to steal a ball that the child is already moving, the discrepant aspect is almost certainly Andy's behaviour, not the child's original action.

3. **Interactivity of aspect:** Groups may combine both aspects detected through direct child action and passive (observational) aspects[8], but only where interactivity does not seem to be central to the child's having detected that aspect as discrepant. For example, *Novel sound effect* (AG8) includes sounds that occur automatically, and those resulting from the child manipulating objects. In all cases the child specifically reacts to the sound. In cases where interactivity does seem central to discrepancy detection, interactive and passive aspects have been kept in separate groups. For example, the child gaining a ball-sorting reward through his own action (AG4), versus observing the reward when it is a result of Andy's action (AG1).

4. **Information loss:** Groups should not be combined (or individual aspects added) if doing so seems like it would lose information, with respect to understanding the current context or making recommendations for future designs.

5. **Theoretical interest:** Aspects or small groups with features that are of special theoretical interest or known developmental importance in the autism literature should be kept separate, so that these features can be analysed further if needed and are not "lost" within a grouping with less theoretical interest.

[8]Discussed in detail in Section 6.2.2.

6.2.1.3 Categorisation results

Grouping aspect labels based on similarity and heuristics produced a list of 19 mutually exclusive AG: 8 for novel aspects and 11 for surprising aspects. The AGs vary substantially in their number of constituent aspect labels, and the number of unique discrepancies represented by that group. Each AG has a brief written description, describing the nature of that category and its members. In many cases, the descriptions refer to related or contrasting AGs, and may also specify why groups of aspects were *not* combined (e.g. theoretical interest, information loss). A selection of definitions, chosen to illustrate some of the variation in the dataset, is reproduced in Table 6.2. Appendix E reports all the aspect group descriptions and lists all constituent labels in each group, along with the number of child reactions to each label. Note that, as all discrepant aspects are definitionally part of a pair with a child reaction, the current descriptions omit repeated references to reactions. The presence of a child reaction in relation to each aspect is assumed.

Table 6.3 lists the 19 aspect groups in relation to DR pair data and participants. For each group it notes the number of constituent aspect labels, total number of unique discrepant instances represented by that AG[9] and the number of children who reacted to something within that AG (i.e. reacted at least once, to one or more of the constituent aspects). For ease of reference, AG are organised alphabetically within their discrepancy subtype.

Several things stand out from this table. First, child reactions are spread unevenly across aspect groups—even keeping in mind the different number of constituent labels making up each group. One AG (13: *Child could not perform an expected object action due to touchscreen or gesture issue*) accounts for 19% of all unique discrepancies in the dataset. This is the AG that captures the substantial difficulties that many children had with the touch-screen interface. The next two largest AGs are *Child discovered novel object action, property, or relationship* (AG6; 11.43%), and *Andy violated a pattern/rule regarding his own usual behaviour* (AG12; 9.21%). Together, these three AGs account for approximately 40% of all unique discrepancies in the ECHOES data set. Outside those three AG, instances are spread quite thinly, which is perhaps unsurprising given the number of aspect labels (112) in relation to the number of total reactions (315). The second striking thing in this table is the variability in the number of children who reacted to each AG. For most aspect groups, only a subset of the chil-

[9]This is not the same as the number of DR pairs, as there may be multiple DR pairs about the same instance, as in the worked annotation example at the end of Chapter 4.

Subtype	Aspect Group	Description
Novelty	**AG2:** Andy demonstrated novel action, object property, or relationship.	The child observed Andy demonstrating a new action or sequence (e.g. jump to shake cloud, pick up a ball and put it in the box), demonstrating a new property of an object (the object itself may or may not already be familiar), or demonstrating a relationship between objects. There is some overlap inherent in this category. For example, demonstrating a new object property may have required a previously unseen action sequence. The duplicative labelling allows for the child's varying focus of attention: s/he may have paid more attention to the object, or to Andy himself. The key point of this category is that the child observed Andy.
	AG4: Child action resulted in ball-sorting activity reward	Child reacted to the sensory reward (bubbles, bees, or fireworks) that appeared when s/he finished correctly filling a box in the ball-sorting activity.
Surprise	**AG12:** Andy violated a pattern or rule regarding his own usual behaviour	Andy behaved differently than would have been predicted from his previously-established behaviour patterns (e.g. delayed entrance, would not accept an offered object). Some instances were passively observed, and the child discovered others by attempting to interact with him. A key point about this AG is that it only includes behaviours that could be considered "general" properties or patterns for Andy, rather than those related to performing a specific activity.
	AG13: Child could not perform an expected object action due to touch screen or gesture issue	The child attempted to use or manipulate an object in a particular way, but did not get the expected outcome due to an issue with the touch-screen hardware or with the child's physical input. As the AG only has non-events in it at this time, the actual outcome is *no outcome* (when some outcome, such as picking up and moving an object, was expected). This AG also includes instances where the child began manipulating an object, but was unable to maintain a continuous, correct touch over time and screen distance (thus dropping it repeatedly).
	AG17: Unexpected or altered property of a familiar object	Child believed that a familiar (known) object was different than expected in some way. This could be in terms of its appearance (e.g. size, colour) or its properties (e.g. action on it produced an unexpected result).

Table 6.2: Descriptions for selected novel and surprising aspect groups.

Subtype	Aspect group	Aspect labels (#)	Children reacting	Unique discrepancies		
				NSR	PRI	Total
Novelty	AG1. Andy action resulted in ball-sorting activity reward	3	2	3	2	5 (1.59%)
	AG2. Andy demonstrated novel action, object property, or relationship	5	7	6	11	17 (5.40%)
	AG3. Andy made a novel utterance	3	5	1	5	6 (1.90%)
	AG4. Child action resulted in ball-sorting activity reward	3	6	7	6	13 (4.13%)
	AG5. Child formed expectations about Andy	3	8	8	12	20 (6.35%)
	AG6. Child discovered novel object action, property, or relationship	15	7	21	15	36 (11.43%)
	AG7. Child was presented with a new scene	7	4	7	6	13 (4.13%)
	AG8. Novel sound effect	5	5	1	11	12 (3.81%)
Surprise	AG9. Andy interfered with the child's turn in an activity	3	5	6	2	8 (2.54%)
Non-events	AG10. Andy was unresponsive to the child's communication attempt	3	2	0	9	9 (2.86%)
Events	AG11. Andy violated a pattern/rule regarding an activity's rules/goals	6	5	6	7	13 (4.13%)
Mixed	AG12. Andy violated a pattern/rule regarding his own usual behaviour	8	7	7	22	29 (9.21%)
Non-events	AG13. Child could not perform an expected object action due to touchscreen or gesture issue	7	8	31	29	60 (19.04%)
Mixed	AG14. Child could not perform an expected object action (non-touchscreen cause)	8	6	11	8	19 (6.03%)
Non-events	AG15. Child perceived that an aspect was missing	6	5	5	6	11 (3.49%)
Mixed	AG16. Mismatch between partner utterance and environment state	5	5	1	7	8 (2.54%)
Events	AG17. Unexpected or altered property of a familiar object	10	6	6	11	17 (5.40%)
Mixed	AG18. Violation of an internal ECHOES environment/activity pattern/rule	9	5	3	12	15 (4.76%)
Events	AG19. Violation of object expectations derived from out-with ECHOES	3	2	2	2	4 (1.27%)
Total		**112**	**--**	**132**	**183**	**315 (100%)**

Table 6.3: Summary of aspect groups for the current ECHOES dataset, listing aspect group features and the numbers of child reactions accounted for by that group. The number of child reactions to each individual label is listed in Appendix E.

dren reacted to *anything of that type* as being discrepant. This is true even of the aspect groups where all children *definitely* had an *opportunity* to encounter those aspects, such as getting a novel reward at the end of the ball-sorting game (AG4) or *almost definitely* had an opportunity, such as Andy making activity mistakes (AG11).

It is also illuminating to present aspect data by participating child, as a useful estimate of the variety within each child's pool of DR pairs. Table 6.4 reports the number of aspect labels and aspect groups for which each child had DR pairs. For each child, it also identifies the aspect group to which each child directed the largest number of his reactions, and the percentage of reactions accounted for by this category. On average, each child reacted to 25 aspect types, and 12 aspect groups. Considering the participants had a mean of 39 unique DR pairs each, this suggests a high variety in the things that children found discrepant. This may be partly explained by the unplanned nature of the events (and non-events) underlying almost all detected surprises. Even those that occurred fairly regularly (such as Andy breaking rules) were still infrequent compared to other types of events.

This arrangement of data highlights that children varied substantially in how far their DR pairs were concentrated in a single aspect group. Differences between the number of aspect groups per child are not particularly striking, but this "market share" of groups is. Compare Ethan, with 50% of his DR pairs contingent on touch-screen non-events (AG13), to Russell, with only 11% of his DR pairs in his largest group *Andy demonstrated novel action, object property, or relationship* (AG2). A child reacting only about a few aspect groups (or with many reactions concentrated in a single group) may have more restricted interests within the environment than a child who reacted to a large, diverse set of aspects. Based on discrepant aspect labelling, a few children appeared to have a strong interest, or focus, or preoccupation, with a certain group of related aspects. The other children presumably had more varied interests, reacting to a bit of this and a bit of that. It is encouraging for future design that even the *most* narrowly-focused child, Ethan, still made half of his reactions to a selection of other discrepant aspects. He did detect discrepancies related directly to ECHOES content (instead of only to the touch interface): his second most popular aspect group was *Andy violates pattern/rule regarding his own usual behaviour* (AG12).

Child	Unique discrepancies (UD)	Aspect labels	Aspect groups	Largest aspect group	% of UD
Anthony	24	24	12	TIE: Violation of an internal ECHOES environment/activity pattern/rule; Andy violated a pattern/rule regarding an activity's rules/goals	16.67 %
Ethan	44	20	10	Child could not perform an expected object action due to touch screen or gesture issue	50.00 %
Hadi	46	35	15	Child discovered novel object action, property, or relationship	21.74 %
Kalil	52	30	15	Child could not perform an expected object action due to touch screen or gesture issue	19.23 %
Lucy	36	27	11	Child discovered novel object action, property, or relationship	38.89 %
Ollie	33	16	10	TIE: Child could not perform an expected object action due to touch screen or gesture issue; Child could not perform an expected object action: non-touch screen cause	24.24 %
Odell	37	23	12	Violation of an internal ECHOES environment/activity pattern/rule	16.22 %
Russell	43	32	16	Andy demonstrated novel action, object property, or relationship	11.63 %
Mean	39.38	25.88	12.63	--	24.83 %
SD	8.72	6.36	2.39	--	13.03 %

Table 6.4: Number of aspect labels and aspect groups represented within each child's unique DR pairs (primary initiations and non-social reactions). This table lists the aspect group responsible for the largest number of each child's DR pairs, and the percentage of his total pairs it represents. Where the aspect group is tied, the listed percentage represents only one group, not both.

6.2.2 Child interaction with discrepant aspects: passive versus active discrepancies

In Study 1 videos, children exhibited striking differences in their ability to inhibit themselves from touching the screen and interacting with ECHOES. Some children, even when repeatedly prompted, simply *could not* wait and watch while Andy took a turn, or while other events took place. Other children were content to watch Andy for long periods, or might spontaneously pause their own action if they noticed something else going on. Andy's reactive planning prioritised feedback and response to the child, triggered through touch input. This meant that he did not give directions or take activity turns if he was already giving feedback, or could detect (via touch input) that the child was taking a turn. For some children, it had to be their turn all the time! Self-inhibition affected, to some degree, what children saw in the environment.

Together, the observed child behaviours and the priorities of the ECHOES AI planner suggest that some children may not have been recorded reacting to certain events (e.g. Andy demonstrating actions or making mistakes) because their constant, hands-on interaction curtailed their possible exposure to those events. For a child unable to self-inhibit and wait, completely focused on her own activity, certain types of environmental events effectively did not exist. Determining whether children were usually involved in discovering discrepancies "hands on" versus watching them unfold might be a valuable proxy for their overall patterns of interaction, and valuable input for planning a future design that could engage a wide range of children.

6.2.2.1 Classifying active and passive discrepancies

Considering activity and passivity of discrepancies captures the basic distinction between watching and doing. An *active discrepancy* is one in which the subjective inconsistency is either created or discovered through the child's hands-on interaction with the environment. *Passive discrepancy* is one that is created and/or discovered without direct child action. It is something that s/he would need to discover through observation, such as watching a sequence of Andy's actions, noticing an object is missing from a game, or hearing a sound effect. Some active discrepancies are more direct than others, in that sometimes the discrepancy is part of an action itself, but sometimes it is in the *result* of a child's action. It is the difference between "I can't pick up this ball, but I should be able to" and "I dropped this ball and *then* it bounced off-screen". To be considered active, a discrepant aspect must somehow include a direct child action,

Interactivity	Aspect label
Active (direct action)	Child tried to perform disabled action: turn flower into a ball *(i.e. disabled in that activity, but available in a similar activity)*
	Andy stole ball from child *(i.e. interfered with child taking a turn)*
Active (action result)	Child action resulted in ball-sorting activity reward: Bubbles
	Child discovered that balls change colour when thrown through cloud
Passive	Andy was inactive for a long period
	Andy stole an already-sorted ball out of the box
	Child perceived that there was a "new garden" setting *(NB: there was not)*
	Novel sound effect: birdsong, as part of garden background sounds

Table 6.5: Examples of active and passive aspect labels. All or most DR pair instances with that aspect label were classified into the category reported here.

or immediately result from a direct action, such as getting a reward in the ball-sorting game. Each of the aspect labels identified in the previous analysis was revisited, to try to classify its instances as being either *active* or *passive*. Some examples of active and passive discrepant aspects are listed in Table 6.5.

6.2.2.2 Interactivity results

Of the 315 unique DR pairs in Study 1, 61.59% (194) were classified as active, and 38.41% (121) were classified as passive. Two of the discrepancy subtypes, novelty and surprising events, show a fairly even split between activity and passivity. Non-events show a notable imbalance (almost 3/4 active), reflecting the extent to which non-events in Study 1 were dominated by child problems with the touch-screen interface. These represented 48% of non-events, and 19% of *all* unique discrepancies. Activity and passivity by discrepancy subtype are reported in Table 6.6.

As anticipated from the observations that inspired this analysis, individual participants varied substantially on the extent to which their detected discrepancies were

Discrepancy subtype	Active	Passive
Novelty	67 (54.92%)	55 (45.08%)
Non-events	95 (73.64%)	34 (26.36%)
Surprise	32 (50.00%)	32 (50.00%)

Table 6.6: Percentage of DR pairs in each discrepancy subtype involving active and passive discrepant aspects

active or passive (see Table 6.7). Overall, slightly more of the group's detected discrepancies appear to be active. Again, this may be partly driven by the very high number of DR pairs concerning touch-screen difficulties. Five children (Anthony, Hadi, Ollie, Odell, and Russell) seem to show fairly even preference for (or perhaps more accurately, tendency to detect) active and passive discrepancies. Ethan and Kalil appear to show a definite tendency toward active discrepancies. For Ethan, at least, this is due to his repeated reactions about touch screen non-events; he appeared more affected and more annoyed by these than any other child. Lucy is the obvious outlier of the group, with nearly 90% of her unique DR pairs related to aspects that she actively discovered. She was one of the children initially observed to seriously struggle with inhibiting herself from touching the screen and other technical equipment: to stay engaged, she needed to be hands-on, all the time. This may be linked to her developmental age, estimated at approximately two years old (see Study 1 participant data). The other participants have estimated developmental ages of three years or older.

Table 6.7 reported the activity and passivity percentages of children's unique DR pairs. Breaking down these numbers by reaction type (Table 6.8) suggests a different picture: that when children encounter discrepancies in different ways, they tend to react in somewhat different ways. This is important, because a future design that uses discrepancy would particularly want to encourage *initiations* due to their developmental importance. On average, only 30% of non-social reactions were about passive discrepancies; this increases to 44% of initiations. Several children (Hadi, Odell) made the vast majority of their initiations about passive aspects, even though they made a larger number of reactions to active discrepancies. This greater proportion of initiations about passive discrepancies is very interesting, especially when we consider that children were free to react in any manner, to any discrepancy. It suggests that there may have been something different about the passive versus active discrepancies in ECHOES, that motivated a slightly different pattern of reactions.

Child	Active	Passive
AM12	11 (45.83%)	13 (54.17%)
EW56	34 (77.27%)	10 (22.73%)
HK53	24 (52.17%)	22 (47.83%)
KL54	33 (63.46%)	19 (36.54%)
LC13	32 (88.89%)	4 (11.11%)
OC52	18 (54.55%)	15 (45.45%)
OM16	20 (54.05%)	17 (45.95%)
RD11	22 (51.16%)	21 (48.84%)
Total	194 (61.59%)	121 (38.41%)
Mean	24.25 (60.92%)	15.15 (39.08%)

Table 6.7: Percentage of child unique discrepancies (primary initiations and non-social reactions) involving active and passive aspects

Several children show opposite patterns for their non-social reactions and initiations. For example, Hadi and Odell have non-socially reacted to predominantly active aspects, but initiated about predominantly passive aspects. Russell and Ollie, who both reacted to more active discrepancies overall, show the opposite pattern of initiating primarily about active aspects and reacting non-socially to passive ones. Lucy's strong preference for active aspects remains across both reaction types.

6.2.3 Between-activity differences in interaction

ECHOES activities[10] afforded different patterns of child-system interaction, due to the nature of the objects involved and the activity goals. Two particular types of difference—the overall flow of interaction, and its degree of structure—are considered here, as these will likely be important for designing future activities. Appendix F presents specific recommendations derived from this discussion.

6.2.3.1 Flowing, starting and stopping

The first time each ECHOES activity was played, children tended to do a certain amount of starting, stopping, and social referencing to the researcher for instructions and reassurance. Andy's initial demonstrations were usually important in communi-

[10] Described in Appendix A.

	Non-social reaction		Primary initiation	
Child	Active	Passive	Active	Passive
Anthony	2 (66.67%)	1 (33.33%)	9 (42.86%)	12 (57.14%)
Ethan	21 (87.50%)	3 (12.50%)	13 (65.00%)	7 (35.00%)
Hadi	21 (67.74%)	10 (32.26%)	3 (20.00%)	12 (80.00%)
Kalil	15 (71.43%)	6 (28.57%)	18 (58.06%)	13 (41.94%)
Lucy	10 (83.33%)	2 (16.67%)	22 (91.67%)	2 (8.33%)
Ollie	4 (44.44%)	5 (55.56%)	14 (58.33%)	10 (41.67%)
Odell	15 (75.00%)	5 (25.00%)	5 (29.41%)	12 (70.59%)
Russell	4 (33.33%)	8 (66.67%)	18 (58.06%)	13 (41.94%)
Total	92 (69.70%)	40 (30.30%)	102 (55.74%)	81 (44.26%)
Mean	11.50 (66.18%)	5.00 (33.82%)	12.75 (52.92%)	10.13 (47.08%)

Table 6.8: Percentage of each child's non-social reactions and primary initiations involving active and passive aspects.

cating what to do, and were often repeated. Once children were confident in an activity (usually from the second or third time onward), two main patterns of play emerged:

1. Activities that supported (or allowed) the child to engage in long, continuous actions, almost entering a flow state;

2. Activities with repeated "units" of action or interaction that added up to meet a goal, usually creating a stop-and-go rhythm of interaction.

The main examples of the former are *Cloud Raining* and *Flower Growing (in flower pots)*. Here, the child can hold down the cloud and drag it around the screen, making rain and growing flowers indefinitely. S/he could grow 50 flowers if desired, or move the cloud across the screen growing a flower in each pot without stopping. Once the child had learned what to do, it was possible for the child to start the games before Andy entered, and complete them without any interaction with Andy or the researcher[11]. Especially with Cloud Raining, children often seemed to enter a flow state, becoming very calm and focused. For this reason, Cloud Raining was a good warm-up game at the beginning of each session, as all children could do it easily and it helped them become focused on ECHOES.

[11]Cloud Raining did not have a set end point; most often the researcher told the child it would be finished when s/he had grown a certain number of flowers, per the on-screen counter.

Activities like *Pot stacking*, *Throwing the ball through the cloud*, and *Ball Sorting* had a very different pattern of interaction because they did not allow continuous activity. Each pot stacked or ball used was one “unit”, adding up toward the goal. Due to difficulties with dragging and lifting objects on the touch-screen, it often took children considerable time and effort to complete each unit. The result was that children did not become as deeply immersed as they did with Cloud Raining, but engaged in much more attention-switching between different on-screen objects, Andy, and the researcher. Completing each unit also provided a natural break for the child or researcher to initiate communication, as well as a mini-achievement to initiate *about*. As they found the tasks somewhat tricky even after multiple plays, children were also much more willing to pause their own efforts and wait for Andy to take a turn, or to ask the researcher for help.

New activities should consider this apparent trade-off between continuous actions that may facilitate engagement and flow, versus repeated units of action that create natural breaks for interaction. This issue is further discussed in Appendix F.

6.2.3.2 Activity structure

In addition to the ECHOES activities that have been repeatedly discussed and referred to in earlier chapters, ECHOES also had a “free play” or sandbox-style activity. It was infrequently used in later sessions. There, multiple familiar objects were available, as was Andy. The original intention with these activities was that children might re-create favourite play from other activities, combine objects in new ways, and—perhaps—invite Andy into their play by giving him objects.

This idea faltered in practice, even for the children who had been more actively exploring ECHOES all along and did not usually wait for instructions. They seemed perplexed by the lack of (or perhaps conflicting) cues, and lack of an obvious goal. The videos show the researchers struggling to explain to children that they could play with any objects they wanted. It was difficult to encourage them without giving specific instructions (i.e. giving the child a goal). It was also difficult to communicate to children that they could obtain (or rather create) favourite objects by taking multiple steps. For example, the child could use the magic cloud to grow a flower, turn this into a ball, and then throw it through the cloud. This was simply too complex, especially in comparison to other ECHOES activities which had clear, single goals. In short, free play was *too* free, and children did not know what to do. Appendix F discusses activity structure further.

6.3 Reflection and synthesis: Learning from discrepancy in ECHOES and considering transfer to new contexts

This section reflects on Study 1 discrepancy-reaction results, and considers them in relation to autism literature. It formulates working hypotheses about the patterns of interaction seen in Study 1, and discusses how these relate to the goal of transfer between ECHOES and new contexts. As noted in the chapter introduction, the positive communications around discrepancy seen in Study 1 are surprising for several reasons, not least that they appear to conflict with literature on the "need for sameness" and restricted and repetitive behaviours (RRB)[12]. Indeed, the literature might have led us to predict the opposite study results: that where children identified discrepancy, they would be uncomfortable, and would find it stressful and disruptive when the environment apparently broke its own rules. Instead, children were most often intrigued and amused. Why might this have happened? Moreover, children repeatedly exhibited relatively rare social behaviours as they reacted to discrepancy, spontaneously initiating to the researcher and Andy. What did they find so compelling—or facilitative—about this situation?

This section poses a *working hypothesis* that would accommodate both the recent findings and the customary understanding of RRB: that ECHOES was a *motivating but manageable* setting and experience, counterbalancing discrepancy with a substantial and "protective" sameness. It uses this hypothesis as the basis for framing the goals of a future design strategy that aims to encourage initiations about *discrepancy-detection opportunities* (DDOs) in new technology contexts.

6.3.1 Reconciling RRB and positive experiences of discrepancy-detection

At first look, Study 1 findings sit uneasily with the "core" ASC characteristic of *restricted and repetitive behaviours* (RRB), sometimes described as a "need for sameness" (NFS; see Chapter 2). RRB may manifest as insistence on specific routines, unusually intense and narrow interests, and experience of changes as highly disruptive. The ASC literature would not predict that children on the spectrum would show

[12] Part of the diagnostic criteria for ASC, as listed in the DSM-5 (2013).

(positive) interest in discrepancies, let alone socially share them with another person. Children's behaviour regarding ECHOES' discrepant aspects suggests that they can potentially find an experience of "lack of sameness" to be interesting and motivating. If not always positive, the experience is at least neutral and not particularly disruptive.

The simplest conclusion is that there is *not* actually a serious conflict between the literature and the Study 1 findings. ECHOES appears to have struck a fortuitous and productive balance between sameness and (unplanned) discrepancy. The VE appears to have violated expectations and/or introduced novel aspects in such a manner as to attract interest and be "worth communicating about", while still keeping enough sameness so as not to be overly confusing or threatening. Observations of children with ASC in the context of ECHOES suggest that there must be a more complex, "grey-scale" picture of how the autistic *need for sameness* (see Chapter 2) actually operates in the world, beyond "same good, different bad." A need for sameness may be fulfilled without *absolute* sameness.

Reconsidering the ECHOES VE and its evaluation context reveals that there was actually *substantial* sameness to balance out the surprises children detected, and the new aspects that ECHOES designers deliberately introduced. Within the VE, all main activities took places within the same Magic Garden setting (background plus ambient garden sounds). Andy was always present, in the same outfit, and with a limited range of expressions and movements. There was a limited set of digital objects, re-used across activities though not always with the same properties (this did create some problems, see Section 6.4.2.3). ECHOES also established strong patterns maintained across activities: the garden scene with objects faded up through black, and after approximately 5 seconds Andy entered from the left. Unless the child took an action first (causing a change in Andy's planning), he would always introduce or demonstrate each activity in some way—not only the first time, but every time. Despite their differences in objects and goals, the activities all had a feeling of sameness due to their look-and-feel, and omnipresent Andy. In some ways, the level of sameness and the clear routines in ECHOES may be what made discrepancy-detection possible at all. Things that were objectively new or different or erroneous were subjectively *noticeable* as such; they stood out from usual play. At least for young child users, such differences may have been far less perceptible in an environment that natively included many more elements, or was governed by more complex, less obvious rules.

The physical setting in which children used ECHOES also provided substantial sameness and routine, which may have played an important role in making their overall experience positive and emotionally manageable. At each evaluation site, ECHOES sessions were always held during school hours in the same room of the building. Sessions always lasted about the same amount of time, and followed generally the same pattern. The same people were present for sessions—either exactly the same, or 2-3 at a time out of a small group of familiar adults. Adding up the sources of sameness within and around ECHOES reveals that there was actually relatively *little* change. New or different aspects were always surrounded by familiar ones.

The key insight here is that *sufficient sameness appears to "balance out" unpredictable aspects*, and when they are thus balanced, children can interact with them in positive ways. In ECHOES, a very large volume of "sameness" counterweighted fairly limited discrepancies. However, it is possible that a closer balance might be possible and still perfectly manageable for most children. This is a question to be addressed by evaluating new designs. Whether this same pattern (of sameness balancing discrepancy) would hold outside of virtual contexts remains to be determined. The fact that these discrepancies are "contained" in a screen, rather than out in the world, may also be important to its manageability—though this is as yet an untested hypothesis.

Sameness alone may not be enough for a manageable experience, however. ECHOES provided another important affordance: first, interaction was *flexible*. Due to other design decisions around facilitating exploration and playfulness, children almost always had multiple options of what to do within an ECHOES activity, and could freely shift their attention between different objects or parts of the screen. They were almost never "forced" to interact with new/different things, either when these were deliberately introduced or occurred due to errors. This may have been important for discrepancies being experienced as manageable and is considered in more detail in Section 6.4.2.3.

The Study 1 findings and circumstances of technology use can be formulated as a *working hypothesis* about child reactions to discrepancy within a technology: that when discrepant aspects are balanced with sufficient sameness within and around a technology, children can experience them as *motivating, but manageable*.[13] Discrepancies were substantial and interesting enough to attract notice and motivate communication, but padded by enough sameness and routine that their presence was not upsetting. Trying to re-create the positive discrepancy-reaction pattern seen in ECHOES (i.e transfer them to a new context) would thus mean trying to create both halves:

[13]A technology-based environment that includes discrepancy in such a way that it motivates positive child interaction is henceforth referred to simply as a *motivating but manageable* environment.

motivation (discrepancy[14]) and manageability (sameness).

The introduction to this chapter stated a goal of creating tools to support *transfer* of the lessons learned in ECHOES to other related contexts. The concept of transferability has been discussed in Chapter 3. The design recommendations and surrounding discussion of context are intended as tools for other researchers and designers to judge the similarity and relevance of the current sending context to potential future receiving contexts that deliberately try to create child communications around discrepancy.

6.3.2 Discrepancy and future design

Any further research questions about creating motivating but manageable environments supervene upon a simpler question: is it even possible to create discrepancies on purpose, so that they *can* be studied?

The nature of discrepancy (as described in Chapter 4) offers a partial answer: we cannot create *discrepancy*, because this phenomenon relies on subjective inconsistency between an individual's experiences and the world. This does not preclude a design strategy that tries to use discrepancy as a tool. By learning from the existing discrepancy examples in ECHOES, it should be possible to identify events and aspects that children will *likely* find discrepant[15]. The designer cannot create discrepancy directly, but can engineer *discrepancy-detection opportunities* (DDOs). This means first providing experiences that will help children establish clear expectations about an environment, and then later *objectively* violating them (or adding objectively un-expectable aspects) in various ways that we think children are likely to find *subjectively* inconsistent (i.e. detect as discrepant).

Before moving on, it is worth considering the type of opportunities that DDOs present to children. What are they an opportunity *for?* The autism literature discussed in Chapter 2 has already reported other instances of researchers or clinicians trying to foster communication by altering the circumstances of a child's routine or physical environment (e.g. Constable, 1983; McClenny et al., 1992; Howlin, 1998; Wetherby and Prizant, 1989; Layton and Watson, 1995). For example, a child discovers that his toy

[14]More accurately, *discrepancy-detection opportunities* (DDOs), see Section 6.3.2.

[15]Throughout this chapter, there is reference to what "children" like, notice, or find discrepant. As explained at more length in Section 6.4.1, the current work continues to be concerned with a sub-population of children with ASC similar to those who participated in ECHOES: in an approximate developmental range of 2-6 years, with some comprehension and production of verbal language (ideally phrase-level language). They might have any SCQ score (or equivalent estimate of autism severity), while still meeting the language criterion. For simplicity, reference to "children" without any modifiers can be assumed to refer to this group of children.

is out of reach or his snack is in a complicated container—the situation "demands" that he communicate, and moreover implicitly suggests a particular type of communication (e.g. request help). Twatchman frames this same idea slightly differently, as "staged problem-solving situations" (1995, p.152). These types of opportunities are all "tangibly maintained" (White et al., 2011), in that communication will secure desired objects or otherwise manage the environment.

There are parallels and differences to the situation observed in ECHOES. In both, changes or inconsistencies are not set as an overt task, but are made available, for the child to find and communicate about if s/he wishes to do so. However, the nature of the communication opportunities is different. ECHOES discrepancies were, in the majority, neither "demands" for specific communications (e.g. show, request), nor problems in the sense of situations requiring a solution. Noticing Andy making a mistake definitely violates a routine, but actually does not impinge on the child at all. Few of the ECHOES discrepancies had tangible rewards for communicating about them, the exception being communication about "stuck" or "broken" objects that elicited researcher help. Even though they were ambiguous (i.e. did not prompt any specific type of communication through the nature of the opportunity) and not tangibly maintained, discrepancies still elicited many child initiations. This suggests that *there is something inherently interesting or inherently satisfying about the process of discrepancy detection,* even when it has no tangible reward and does not solve a problem.

To answer the question posed earlier, the *opportunity* in discrepancy-detection opportunities is, for the present, understood in the sense of the existing ECHOES examples: an ambiguous situation about which many types of communication or action may be relevant. The opportunity is not a demand for a specific action, and with the exception of requests for help, will have no reinforcement via tangible (i.e. "real world", non-digital) results. Like both ECHOES examples and literature "demand" examples, the opportunities are embedded in a familiar environment without announcement and *not* set as an overt problem-solving task. Discussion of the nature of DDOs is resumed in relation to the high-level design principles in Section 6.4.2.3.

6.4 Recommendations for a future design

Previous sections of this chapter have further analysed the details of ECHOES DR pairs, and have attempted to distil the most important ingredients of what has "gone right" in ECHOES with respect to balancing DDOs and predictability. The current section groups this distilled information around six core concepts, which represent preliminary, high-level principles for manageably incorporating DDOs into future interactive technologies. Additional guidance for designing young children's ASC-tech, based on "lessons learned" from ECHOES, is included in Appendix F.

The design guidance in this chapter is best understood as one side of a transferability dialogue with the reader. The taxonomy development and study results have given an account of the sending context of the research, enabling readers to make similarity or relevance judgements about the "fittingness" between this context and their own potential receiving contexts (Lincoln and Guba, 1985, p.124). The original author cannot know about every possible context to which other researchers may wish to transfer the work, and thus cannot judge where transfer is appropriate and to what degree. Only the *reader*, with a detailed knowledge of the original research context *and* the potential new context, can make a judgement about "how far they can be confident in transferring to other situations the results and conclusions presented" in a piece of research (Shenton, 2004, p.70). The recommendations are meant to aid the receiving researcher in understanding which factors in the current context appear to have affected the phenomenon of interest (discrepancy), and in what ways. They try to identify which circumstances might need to hold in a new context and which elements could potentially be transferred in order to foster similar discrepancy-related communications. A crucial part of this transfer will be the replacement or expansion of the "naturally-occurring" discrepancy detection opportunities in ECHOES with deliberately designed ones. How to strategically choose and manageably include these opportunities is a main concern of the guidance.

6.4.1 Assumptions about future design

The recommendations in this chapter make a number of starting assumptions about the context, users, and overall goals of a future design. Overall, these assumptions serve to focus the recommendations on circumstances similar to the original ECHOES system and evaluation. At this early stage of investigation where discrepancy is still poorly understood, this makes sense because it is not yet clear how far the phenomenon might extend with respect to other participant groups, and how much the ECHOES philosophy and context contributed to the overall positive outcomes of children initiating about discrepancy.

The recommendations assume participants or users with a similar profile to those in Study 1: developmentally 2-6 years, with approximately phrase-level language production. They may have a range of SCQ scores (i.e. range of severity/presentation of autistic characteristics). The recommendations also assume that the overall design goals, environment structure, and child experience are broadly similar to the ECHOES system and evaluation. Regarding the **technology contents and structure**, the recommendations assume:

- The design is intended for individual rather cooperative play;
- The content is planned for presentation/use over multiple sessions, to allow sufficient exposure to the environment;
- There are one or more self-contained activities with identifiable goals;
- That both the environment and any individual activities:
  - are game-like in the sense of facilitating play, but are not competitive;
  - are oriented toward interaction and exploration rather than performance of specific skills;
  - may embed skills practice of some kind, but do not use "drill-and-practice" type activity;
  - allow children to work/play at their own pace;
  - do not rely heavily on language in order to understand or complete tasks.

The design guidance makes no assumptions about how extensive a future design might be, how many activities or sub-parts it might have, and how many different skills are targeted (if any). Note that there is also no assumption of a virtual character, or any characters[16].

Regarding the **physical environment around the technology and its overall manner of use**, recommendations assume that the environment:

- Is already at least somewhat familiar to the child and stays the same across most or all sessions (in terms of what and who is present);
- Has a routine around the technology use (i.e. how session starts and ends, how the technology is introduced/explained...);

[16]The additional design recommendations (Appendix F) discusses character behaviour, but overall the design guidance treats character(s) as one possible design option among many.

- Has an adult continuously available to the child during technology use, able to respond or assist where necessary.

There is also an assumption that the child will be the one to primarily act on, and have perceived control over, the virtual environment. Adult partners are assumed to be largely reactive during the main use (e.g. once the environment is set up, child has received any instructions, etc). Thus far, the design guidance makes no specific assumptions about the exact context of use (e.g. home, school, therapeutic or clinical) or who the adult partner is.

These recommendations also make a central assumption about the **goal of including DDOs**: that the end goal is to encourage children's spontaneous communication with a social partner, particularly initiations. This means that children's discrepancy-detection is not considered inherently valuable, but valuable as a means to an end. DDOs are of value where they motivate subsequent communication. Relatedly, the recommendations assume that **discrepancy-detection will not be treated as a task in itself**: the goal of an activity is never to find the mistakes, or find what is new or different. It is very possible that a design that challenges children to find mistakes might be very successful at motivating communication. However, exploring that possibility is reserved for future work.

Finally, the recommendations assume that there will be a **high value on ensuring that participants have a positive experience** with the technology. This may mean managing difficulty and cognitive demands, adding rewards and positive feedback, and making decisions about how the adult social partner should intervene (or not) to support the child. This value on positive experience is about more than making sure children return for a second or Nth technology session. Keay-Bright and Howarth write (in relation to ASC-tech) that

> ...creating environments whereby people feel good about themselves, where they feel able, where they are not prejudged, is a vital prerequisite if we are to support learning and social communication....computers offer huge potential to assist in managing complexities, enhancing creativity and enjoying social communication. Therefore, it is imperative that we approach the design of technologies in the spirit of assisting people in gaining independence by feeling good about themselves, so that they are empowered to make a positive difference in their own lives. (2012, p.139)

For people with autism—and perhaps any disability—positive and successful interactions with technology may have inherent value and benefit well beyond a design's intended goals.

6.4.2 High-level design principles

6.4.2.1 Overview

These principles aim to help designers motivate child initiations via DDOs and to ensure that children's overall system experience is positive and emotionally manageable. While it is general good practice in HCI to design systems and interactions that are coherent, comprehensible, and largely predictable to users, the current goal is more complex: to create a system coherent and predictable enough that its behaviour can be a talking point rather than a threat for children with ASC—even as it repeatedly violates their expectations. The task is thus to target the fruitful middle zone between complete sameness and unmanageable chaos.

The author offers six high-level principles intended to guide transfer of the *motivating but manageable* pattern of interaction seen in ECHOES to future technology designs. Each principle describes a factor thought to be important in terms of contributing to the overall effect of a design which can facilitate positive discrepancy-reaction activity. In effect, each is its own working sub-hypothesis, under the larger working hypothesis of motivation and manageability. As a summary, it is recommended that designs attempting to create a motivating but manageable environment for discrepancy-contingent communication should:

1. Maintain high *integrity* in the environment and in individual activities, such that "usual" rules and relationships are clearly established early in child-environment interaction, and largely maintained over time. **INTEGRITY**

2. Allow *flexibility* of interaction within activities, such that children are never "forced" to interact with a DDO that they may find uninteresting, incomprehensible, or stressful. **FLEXIBILITY**

3. Ensure that DDOs that alter or prevent customary actions are always *resolvable*, and do not create unachievable goals or the perception of the environment being completely "broken". **RESOLVABILITY**

4. Offer a wide *variety* of DDOs to accommodate child interests and needs. **VARIETY**

5. Include DDOs at a *frequency* that is regular, but will not make them a major proportion of total child-system interaction. **FREQUENCY**

6. Offer DDOs that are *ambiguous* communicative motivators, rather than DDOs that make implicit demands for the child to initiate using specific behaviours or for specific purposes. **AMBIGUITY**

These principles were formulated by working across multiple sources of information, including Study 1 results and "lessons learned" across the ECHOES project, autism literature, and logical implications of the discrepancy concept. The following section traces some of the most important contributions to each principle, and section 6.4.2.3 lays out each principle in more detail.

6.4.2.2 Information sources contributing to principles

The high-level design principles summarised above each draw on multiple sources of information. In most cases, these sources reinforced each other, but in a few cases suggested different conclusions. This section gives a name and brief description of five main information sources. Below each source is a selection of key contributions to the principles in the following sections. This is intended to be an overview of the nature of the specific contributions each source made, rather than a comprehensive account.

Logical implications of a discrepancy-detection strategy: The definitions of discrepancies and child reactions logically imply that certain conditions must be in place in order for discrepancy-detection to be possible. For example, in order for a designer to try to violate child expectations, children must have been given an opportunity to establish expectations in the first place.

- INTEGRITY
  - In order to violate child expectations, children must be allowed to establish expectations about the environment, individual activities, and the results of his/her own actions.
  - DDOs must be detectable as new or different in some way; this requires substantial sameness in the environment as a point of comparison, rather than actual/apparent change everywhere, all the time. It also requires that DDOs are "different enough" in type and magnitude to be noticeable by young children.
- FLEXIBILITY: Designing *opportunities* for discrepancy detection, rather than *demands*, suggests that the child should have a viable choice not to interact with

a DDO. If s/he does not do so, s/he should still be able to continue some form of interesting or productive activity.

- RESOLVABILITY: If designing *opportunities* to react/interact, the activity and environment should continue to function if a child does not take up (or understand) a particular DDO. Opportunities may temporarily block a child's progress, but should be able to "self-resolve" if s/he does not act.
- FREQUENCY: All environment elements will be novel when first encountered (i.e. no expectations, unpredictable). The designer must account for this *inherent novelty* in their planning, and estimation of sameness in the environment.

Quantitative results from Study 1: This source includes numerical data from the ECHOES video data study, including affect-related data, counts of discrepancies in different aspect groups, and active/passive classifications. This data is particularly useful for giving a picture of the amount and composition of the children's reactions to discrepancy, and identifying overall trends.

- RESOLVABILITY: All negative child reactions in Study 1 were somehow related to lack of resolution, e.g. child could not complete an action and meet a goal, Andy interfered with child turn (and no way to fix or re-do this).
- VARIETY
  - Many different discrepant aspects represented in Study 1, even though ECHOES designers re-used many objects and behaviours across activities (i.e. high variety).
  - It appears "safe" to include a wide range of discrepant aspects. Children appeared to ignore or react neutrally to discrepancies that did not strongly interest them.
- FREQUENCY:
  - Quantitative summary data indicates how many DR pairs are in the current data and how often they appear. This allows a preliminary estimation of a manageable number of discrepancy-reaction pairs for future designs.
  - Some children were far more reactive to discrepancy, and to ECHOES overall (both socially and non-socially) than were their peers. There was a range of 38-70 DR pairs within a 45 min video sample.

  - A sizeable number of Study 1 DR pairs were *not* related to objectively new, different, or erroneous aspects: for example, children talking about the "new" garden setting or "giant" flowers. These could be considered *perceived* discrepancies. Any estimate of DDO frequency should take into account that there will be additional perceived discrepancies.

Child information: This source considers differences between participating children, drawn partly from Study 1 data, but more from researcher observations and interaction examples. That children interacted with and understood ECHOES in different ways is immediately apparent from watching their videos, but these differences are not always captured well by quantitative data. For example, there is the issue of self-inhibition (see Section 6.2.2), or that some children were eager to try things out in ECHOES, while others stuck closely to what they were shown. In some sense, individual children can be considered almost in the manner of personae for future design (e.g. do we think Lucy would find something to interest her here?).

- INTEGRITY
  - Timing and ordering of environmental events were highly salient to some children. These should be considered as components of overall integrity.
  - Children varied in the time taken to grasp cause-and-effect relationships between their actions and results in the environment. They should be allowed plenty of time to grasp basic activity actions, before relationships are changed for DDOs.
- RESOLVABILITY: Most children perseverated in the face of perceived-broken objects or actions, trying several times to make them work, or to solicit adult help. At least one child (Ethan) was very perseverative, trying the same actions for several minutes, and refusing to be re-directed within the environment. He seemed unable to tolerate the lack of resolution.
- VARIETY
  - Need to include a wide range of discrepancies if designing for a group of children: Low overlap between most children regarding the set of aspects to which they reacted.

  - Children require fairly good self-inhibition and waiting ability in order to be able to observe passive discrepancies such as VC mistakes. Not all children are ready to do this; some reacted almost exclusively to active discrepancies (e.g. Lucy).
  - Some children (e.g. Kalil) actively explored an environment and tried things to see what would happen. Others were less willing to do this, tended to follow VC or researcher-modelled actions.

- AMBIGUITY: Children showed high variability in the form/function of their communications about the same aspect types. This suggests current ECHOES discrepancies will be a good model for future ambiguous DDOs.

Lessons learned from ECHOES project formative and summative evaluation: The ECHOES project team, including the author, spent extensive time working with children with ASC in schools around the UK and in the lab. Observing and collecting data about children's interactions with several generations of the ECHOES system has provided valuable insight into what is feasible with this age and ability group—for example, how long an "average" session with the technology can reasonably be expected to last, or which touch-screen actions are difficult to manage. For designers, perhaps some of the most valuable "lessons learned" relate to the gaps we have observed between intended object use or activity goals and children's sense-making.

- INTEGRITY

  - Sameness external to the technology (session routines, room where used, familiar researchers/staff) may play an important role in establishing a level of overall sameness to be balanced against the discrepancies within the environment (presented via technology).
  - In ECHOES, children sometimes seemed confused by similar activities. Designers should consider activity-level integrity, and give clear cues so that children can discriminate between different activities, and their respective goals and affordances.
  - Children repeatedly attempting disabled object actions (and rejecting, or not understanding, researcher explanations) suggests that objects should have the same properties in each activity where they appear.

- FLEXIBILITY

- Many children spontaneously switched attention to other objects or parts of an activity when faced with something that was perceived as missing or broken; this could be considered a successful form of emotional self-regulation.
- Perception of action/object options and ability to move between them (such as repeated "units" of action with different objects) may be an important factor in manageability, also creates natural breaks in which interaction may occur.

- RESOLVABILITY
  - Many "broken" aspects about which children perseverated were due to errors or disabled object actions and could not be resolved within the environment (or GUI). These have potential for children to proceed from frustration to meltdown, though this luckily did not occur in ECHOES.
  - Presence of researcher GUI with on-demand character instructions (displayed on secondary monitor) was helpful to resolve difficult situations, re-direct a child's attention ("Let's see if Andy will take a turn!")
  - Completing an activity, or sometimes even 1 unit of action, (e.g. sort ball after long period of difficulty) was usually enough to diffuse child frustration about things that were not working.
- FREQUENCY
  - Average ECHOES length session length was 12-15 min and included multiple activities. Each individual activity took children only a few minutes, unless there were many touch screen problems.
  - There will be unavoidable and unplanned errors within a system, including interface difficulties. The designer should keep these in mind as part of the overall level of change and difference within the system (that must be balanced with sameness).

Autism literature, theory and practice: Aside from providing further information about the needs and strengths of this user group, findings generated by decades of autism research and educational practices provide useful big-picture information about what has, and has not, been successful elsewhere. Longitudinal studies also provide valuable clues about which types of child behaviours seem to support positive long-term outcomes, and thus might be most important to support through technologies.

- INTEGRITY
  - *Need for sameness* is a core characteristic of autism spectrum and there is extensive practice-based advice on how to meet this need in daily life, so as to create manageable environments and experiences (e.g. signpost and support transitions, supports in environment such as visual schedules)
  - Difficulties with *generalisation* in autism suggest that expectations may not transfer between activities that a child perceives to be different (i.e. as separate contexts). Designer will need to address this in some way.
- FLEXIBILITY: Transactional support strategies (e.g. SCERTS; Prizant et al., 2005) are a possible model of effective interaction and support that involves adult social partners following a child's interests and attention in a situation. Building on child interests and engagement can be a powerful basis for structuring/motivating communication and social activity.
- RESOLVABILITY: Repetitive or perseverative behaviours are prevalent on the autism spectrum and are considered related to the need for sameness; the pattern observed in ECHOES should be expected elsewhere, and designers should plan how to mitigate it.
- AMBIGUITY
  - Previous face-to-face interventions (and anecdotal clinical reports) that embed communicative demands in the environment give a point of reference for what has been previously successful at attracting attention and motivating communication, and provides a point of comparison for how the current work aims to design opportunities rather than specific demands.
  - Autism literature highlights social initiations (as opposed to imperative initiations) as being particularly important developmentally, and not well targeted by existing interventions (see Chapter 2).
  - Participatory design literature (e.g. Frauenberger et al., 2013) highlights difficulty of adults understanding viewpoint, experience of children with ASC. Embedding demands imposes adult viewpoint about what a situation

means, appropriate actions to take. Ambiguity better allows child to make own meaning, and gives adult opportunity to better understand child.

6.4.2.3 Initial principles

The following six principles (integrity, flexibility, resolvability, variety, frequency, and ambiguity) together attempt to delineate the conditions of an emotionally manageable environment, in which children are motivated to communicate about discrepancies. Such a context is an *overall effect* of the principles. Where possible, the author includes specific examples and recommendations of how these principles might be instantiated in a design. As their purpose is to support transfer, the principles extrapolate away from the current context, and refer to the inclusion of designed DDOs, rather than the "naturally occurring" DDOs in ECHOES.

Note on terms: References to aspects, elements, events, or non-events are all used in the same way as they are defined earlier in this thesis. Reference to activities or games is intended to capture sub-parts of a larger system or technology, in the way that the ECHOES environment is made up of separate activities that take place in the same setting and share many elements.

INTEGRITY An environment with high integrity maintains a level of consistency that makes it possible for a child to determine what is new, different, or missing (i.e. to *detect discrepancies*). For this user group, the required consistency is high. Most aspects should be held constant most of the time. The environment should appear to be a lawful place with occasional changes and mistakes, not one with constantly changing rules or no rules at all. The child should be able to explore, and to develop clear expectations about the content, patterns and functions of the environment (i.e. to establish a body of expectations and knowledge against which to compare later experiences). Forming expectations means establishing how and when one's actions affect the environment and how environmental aspects affect and relate to one another. S/he may need to play each activity at least once in an "original" or "baseline" form to develop these expectations, before surprises are introduced. For children in this young age group, cause-and-effect relationships must be clear and reasonably simple, with a 1:1 correspondence between actions and results. Ideally, results should be immediate, or minimally delayed.

Systems with multiple activities must consider *activity-level integrity*, to make sure that children can tell activities apart, understand which "rules" may apply, and be sure of their current goal. This recommendation is emphasised because it was a weakness of ECHOES, which did not adequately signal different activities in a consistent, language-free way. A single garden background, plus re-use of objects to "keep things simple", meant that all activities looked similar, despite differing goals and object properties. This created confusion when children tried to perform object actions that "should work" only to find that the objects were "broken" (i.e. actions were available in some activities but not others), or looked for objects that "should" be there, but were not a part of the activity. ECHOES participants initiated about these expectation-violations insistently and calmly, but the potential for these situations to create frustration and negative affect remains. Activities with different goals and affordances should be easily differentiable *from the child's viewpoint*—not only by name. Where activities are related (e.g. activities with same core goal, varying in difficulty) this should also be apparent from the child's viewpoint. Designers should not choose DDOs that affect activity-identifying features and introduce confusion about which activity is being played (e.g. do not remove a "signalling" piece of scenery as a potential surprise). Also, DDOs should not fundamentally alter, or prevent attainment of, core activity goals. When activities are very simple with limited objects and actions, an altered core goal effectively creates a new activity.

While activities within a system should be differentiable, it can create new challenges if they are too different. For example, if a technology comprised a suite of stand-alone games (with no common characters, elements, or patterns), each game would effectively function as its *own* environment, needing to establish its own integrity. If there are multiple activities (games, tasks, sub-parts) within a system, their content and rules must be recognisably related (from a child point of view), in order to be treated as part of the same environment (i.e. with respect to defining discrepancy and its subtypes).

FLEXIBILITY In ECHOES, it was rare for multiple parts of an activity or the environment to be concurrently affected by a discrepancy. Novel aspects were always presented alongside familiar ones. AI planner or graphical rendering errors generally only disrupted one object type at a time (or Andy), leaving the rest of an activity "intact" and operating as usual. Children who did not wish to engage with particular aspects could act on unaffected (or less-affected) ones. Those encountering (objectively) new or different aspects could switch their attention elsewhere, continuing the overall flow of play. This flexibility of action (i.e. never being "forced" to engage with a DDO) may be a central reason why ECHOES discrepancies were manageable. The opportunity to detect any given discrepancy almost always lay with the child, as did the option to share it, laugh at it, or ignore it.

Based on these observations, designers are advised to enable flexibility of action by including DDOs that affect only part of the environment at a time, rather than having global effects, and by ensuring that there is always more than one possible action or focus of attention available in the environment so that the child can switch between them. A flexible "order of operations" is one way to do this, e.g. by making activities modular with repeated "units" of action, rather than using unique actions or actions that must be completed in a set sequence. Another option may be to have limited, goal-irrelevant objects or actions available as an alternative, especially should the child become frustrated with the main part of the activity. The challenge here is to choose objects or actions that are interesting enough to investigate, but ultimately less interesting than the main activity.

RESOLVABILITY This principle requires DDOs that alter or prevent customary actions to be resolvable within the environment, and to avoid impasses in which activity goals are unachievable or the environment appears "broken" and unresponsive. Resolving a discrepancy may include direct child action, or may require adult or VC intervention (e.g. demonstrating the correct action, giving a command through a system control panel). For discrepancies directly related to activity goals, it may be wise to make them "self-resolving", in case a child does not notice them, does not understand, or cannot fix them. For example, a DDO might be linked to a timer, so after an interval it will return to its usual state (e.g. appear, become responsive, resume its usual behaviour). Re-starting an activity or the system should also automatically resolve or reset altered aspects (i.e. they should not persist across uses). DDOs that halt the flow of activity and demand a specific action or adult intervention should be relatively rare, as genuine system malfunctions may already provide events of that type!

Finally, note that not all discrepancies require resolution. It does not make sense to talk about resolving a new (potentially novel) aspect, nor does it make sense in relation to DDOs that have no possibility for interaction (e.g. removing a piece of background scenery, removing or changing a sound effect).

VARIETY When designing for young children with ASC in general, ECHOES findings suggest incorporating a wide variety of DDOs, including "something for everyone", whether a child needs hands-on interaction, notices novelty but not surprises, or only likes sound effects. Each child can detect and react to some opportunities, meaning that the overall design strategy can still be of benefit even when nothing is known about the specific end-users, and when (in practice) individuals' patterns of interest and activity are quite different. This recommendation stems from the ECHOES observation that particular discrepancies of great interest to some children were completely unnoticed by others. There was strikingly little overlap between children in the sets of discrepancies they detected, plus wide variation in the breadth of their interests (a range of 16-33 aspect types detected; range of 10-17 aspect groups).

The lack of overlap would be highly problematic if children had reacted positively to "favourite" discrepancies and been upset by others. This was not the case: they laughed and shared excitement about favourites, with visible but more neutral reactions elsewhere. The low percentage of negative reactions to discrepancy may be related to flexibility of interaction. In ECHOES, children almost always had options within the VE, and were not "forced" to interact with or resolve specific (potentially uninteresting or threatening) aspects. Without this flexibility, negative reactions might increase.

FREQUENCY of DDOs As noted in relation to the variety principle, there was relatively little overlap in the specific aspects children detected as discrepant—what fascinated some was apparently unremarkable or even invisible to others. The highly subjective nature of discrepancy means that designers can include a fairly high number of DDOs without worrying about creating chaos—only some of them will "exist" to any one child. Current ECHOES videos can help us to estimate how many DDOs to use in the future. In Study 1, the participants detected 315 unique discrepancies (64 surprising events, 129 non-events, and 122 novel aspects) across 347 minutes of video. On average, this is a surprising event detected every 5.4 minutes, a non-event detected every 2.7 minutes, and a novel aspect every 2.8 minutes. Based on these figures, the design for a 12-minute technology session could likely offer *an absolute minimum* of 4-5 non-event opportunities[17], another 4-5 opportunities for novelty, and at least 3-4 surprising event opportunities. "At minimum" is particularly emphasized because it is highly unlikely that a child would take up all opportunities (see *VARIETY*, above). The current numbers are based only on actual ECHOES DR pairs, and thus are presumably much lower than the number of situations to which children could *potentially* have re-

[17] I.e. DDOs that the designer thinks the child is likely to detect, and in which the child-environment relationship would be expected to constitute a non-event.

acted. Designers need to include enough DDOs, with enough variety, that every child is likely to be interested in a subset of them. Until discrepancy is more precisely understood, it might be wise for designers to include many DDOs—with high variety—in the expectation than any individual child would react to only a minority of these.

AMBIGUITY A potential benefit of DDOs as a design strategy is that they appear to facilitate interesting and open-ended communicative opportunities, to which the child can react for any purpose, using a range of behaviours. For example, after observing Andy make a mistake (e.g. as in worked annotation example, Chapter 4) it is equally relevant to initiate by correcting him, by commenting and pointing, or by sharing laughter with the researcher. The alternative to such ambiguous motivators is embedding specific communicative demands, (e.g. an object placed out of reach "demands" a help-seeking initiation; see Constable, 1983; Howlin, 1998; McClenny et al., 1992; Twatchman, 1995). The *FLEXIBILITY* principle grants children some freedom of action within an environment. Ambiguity seeks to enable freedom of *re*action. In facilitating initiation, designed ambiguous opportunities seem more similar to "daily life" communication opportunities than are very specific demands, possibly forming a better basis for skill generalisation (as yet an untested hypothesis).

From a research standpoint, ambiguous DDOs provide a valuable window into the interests and attentional focus of young children with ASC, and can illuminate the often-significant gaps between an adult designer's intentions and the child's experience of an environment. Ambiguous opportunities (as opposed to demands) may also allow a design to be relevant for a wider range of children, because specific skills are not required in order to engage. Children can meet the environment and social partners where they are, in terms of their current communication behaviours.

6.5 Summary

This chapter has reported three linked pieces of work that construct a bridge between discrepancy in the ECHOES context, and future designs through which to investigate discrepancy in a new context. The main data from in Study 1 did not include the actual aspects that children found discrepant. The descriptive data collected during annotation was labelled and categorised here, producing 112 discrepant *aspect types*, grouped by similarity into 19 mutually exclusive *aspect groups*. Viewing this data by child reveals that there was very little overlap between individuals in terms of what they found discrepant. Children also varied in the extent to which their reactions were concentrated in a small number of types and groups, versus widely dispersed. Aspects were additionally classified as being detected through a child's direct activity, or through passive experience (*active* versus *passive* discrepancies). Across the group, 61% of DR pairs involved active discrepancies—a number slightly skewed by two children who overwhelmingly reacted to active discrepancies, with the other children showing a more even split. Considering initiations only, some children appeared more likely to initiate (than non-socially react) about passive discrepancies. A child's ability to self-inhibit behaviour and wait are hypothesised to be important determinants of the extent to which s/he will encounter and react to passive discrepancies. The pattern in ECHOES suggests a future designer should definitely include both types, remaining aware that active DDOs, especially those related to the main activity actions and goal, are likely to be encountered by a greater share of participants.

The apparent contradiction between the positive, discrepancy-related interactions in Study 1 findings and the autistic *need for sameness* can be resolved by the working hypothesis that sufficient sameness is able to "balance out" discrepant aspects. When this balance is present, as it appeared to be in the ECHOES context, occasional discrepancies can interest and motivate children, rather than upsetting them. The environment *as a whole* remains stable, predictable, and safe. Sameness outside a technology—such as the physical context of use and familiar supporting adults—may be important to a child's overall experience of sameness. This concept of a *motivating but manageable* context of interaction forms a high-level blueprint for future design. Combining this idea with an examination of the discrepancy concept identifies what it is that we are trying to include within a motivating but manageable environment: *discrepancy-detection opportunities* (DDOs). Designers cannot directly create the subjective phenomenon of discrepancy, but *can* create situations in which children appear likely to detect a discrepancy, based on extrapolation from ECHOES findings (e.g. violating a previously-established pattern, introducing an objectively new object).

The working hypotheses developed around discrepancy are meant to guide transfer of ECHOES findings and insights (i.e. a deeply understood *sending context)* to a new *receiving* context. While similarity and relevance judgements ("fittingness" judgements; Lincoln and Guba, 1985) are usually left to the reader, overt design guidance can support this process. Six high-level design principles, based on a range of empirical, literature, and experiential sources, draw attention to factors in the current context that appear to have affected the phenomenon of interest (DR pairs), and in what ways.

They try to identify which circumstances might need to hold in a new context and which elements should be transferred in order to foster a similar discrepancy-reaction pattern. The design principles look ahead to a new designer's task of choosing and placing designed DDOs that will replace or expand the "naturally-occurring" DDOs in ECHOES. How to strategically and manageably include these opportunities is a main concern of the guidance, and of any future design in this area. Additional guidance based on ECHOES "lessons learned" supplements the content in this chapter (see Appendix F). This comments on specific choices a future designer might make to successfully engage and support young children with ASC, regarding activity structure, virtual characters, touch-screen actions and similar.

Chapter 7 reports the development of a new design (*Andy's Garden* game suite) which aims to create motivating but manageable child communication about discrepancy, guided by the principles reported in this chapter. An evaluation study of *Andy's Garden* is reported in Chapter 8, to determine whether the design has been able to broadly replicate the pattern of DR pairs and provide proof-of-principle that DDOs can motivate communication in a new setting.

Chapter 7

Developing new touch-screen games with discrepancy-detection opportunities

7.1 Introduction

Study 1 raised more questions about discrepancy than it was able to (provisionally) answer. They centre on whether it is possible to create an emotionally manageable environment that motivates child communication through *discrepancy-detection opportunities* (DDOs). Chapter 6 proposed working hypotheses about which factors and circumstances in ECHOES enabled the original pattern of interaction, and design guidance to support transfer of this pattern to a new context. Rather than transferring insights from Study 1 into an existing, similar context that we wish to affect, *we are using those insights to create a custom receiving context intended to maximise transfer*, i.e. designing a new set of games. At this point, when little is known about discrepancy and how children experience it, the best strategy is to hold fairly close to the original ECHOES environment—elaborating on its perceived strengths, but not slavishly re-implementing it. A new design can follow up on some of the leads generated by Study 1, actively expanding the picture of discrepancy in addition to trying to re-create some of the original findings. Indeed, a better understanding of discrepancy is perhaps the highest-level goal for a new design. As already discussed earlier in the thesis, the current work does not seek to create an intervention that changes children's communication behaviour outwith the environment. It is interested in using design as a tool to investigate—and potentially affect—children's behaviour during game play. All questions about behaviour in other contexts must be reserved for future work.

Re-creating an ECHOES-like pattern of discrepancy-contingent interaction would more specifically mean:

1. Including designed events and situations (DDOs) that children notice, and react to as being discrepant in some way (DR pairs);

2. That all children in a group detect and react to some discrepancies, designed or otherwise, but may differ widely in the specific things to which they react;

3. That a sizeable portion of reactions to discrepancy (DR pairs) are initiations, directed to a human or virtual social partner;

4. That children's overall experience appears to be positive, and emotionally manageable.

Three mini-games have been developed as a tool to investigate discrepancy: Flower-Growing, Apple Sorting, and Carrot Growing (very similar to flowers). Together with menus and introductory material, they make up the *Andy's Garden* environment. Games draw from the original ECHOES "Magic Garden" setting, which includes cause-and-effect play, and an exploratory, non-competitive format (originally developed with extensive stakeholder input). They include a virtual character (VC) with the same image and a similar helpful role as the original Andy. Each game has a "basic" version to introduce the child to the environment's usual contents, patterns, and behaviours. In later games sessions, the child will play two different "altered" game versions, that include a range of DDOs that introduce objectively novel or changed elements, intended to facilitate novel, surprising, and non-event DR pairs. DDOs are expected to interest children, based on the interactions in Study 1, and pose opportunities for them to spontaneously initiate communications with a range of goals and behaviours. Evaluating the games (Chapter 8) will be a test of the working hypothesis that a motivating but manageable environment with embedded DDOs (as described in Chapter 6) should create an interaction pattern similar to that of Study 1.

This chapter begins with some of the constraints on the designs, imposed by the model of ECHOES and by the projected evaluation context. The game evolution is described and visually illustrated, from early concepts through to semi-final versions, which were usability tested with typically developing children. Very important to the final designs is the selection of DDOs, and their arrangement across the mini-games and different days of game sessions. The chapter reports several heuristics developed to manage the process of placing designed DDOs in a way that respects activity integrity. As the current work has used an existing game engine (GameSalad Creator) rather than custom software, details of technical development are kept brief. The chapter concludes by stepping back to consider how the design choices have sought to reflect the six principles for motivating but manageable design (introduced in Chapter 6).

Note on chapter terminology The following terminology is used in the current and subsequent chapters to refer to pieces of the new design:

Environment Refers to the content and behaviour of *Andy's Garden* as a whole (the three mini-games and surrounding material such as menus and open/close sequences).

Game Used interchangeably with environment.

Mini-game A small, self-contained, interactive activity with a goal, a constituent part of *Andy's Garden*.

Activity Used interchangeably with mini-game; used in the same sense that this term was used for the original ECHOES.

Basic game versions The "baseline" versions of the three mini-games, used to establish expectations about normal behaviours, patterns and actions. These games *do* still provide opportunities for discrepancy-detection, because their elements will be objectively novel to the child player (i.e. *inherent novelty*). These are the games children will see in their first session with *Andy's Garden*.

Altered game versions These mini-games are the same as the basic versions, except that they now introduce DDOs, which affect some game elements, some of the time. Two different altered game versions have been developed, for children to play in their second and third sessions with *Andy's Garden*.

7.2 Game development overview

7.2.1 Stages of development

The game development process was relatively linear; it did not go through many cycles of user testing and revision. It also did not engage in a broad exploration of the problem space, generating many initial ideas and then narrowing down to the concepts that were eventually developed. As discussed in section 7.3.1, a number of initial decisions were made to substantially constrain the design process—namely, to build fairly directly on ECHOES activity concepts and mechanics.

Two game concepts (growing and sorting; section 7.3) were implemented in the GameSalad game engine (described in the next section) very much as initially envisioned. These basic versions were developed to substantial completeness, at which point they were critiqued by HCI and children's technology experts (section 7.4.1). The remainder of the basic game development incorporated this input. Designing DDOs was not undertaken until the basic game versions were complete, and thus all "baseline" contents and behaviours were finalised. These would be the basis of the child's expectations, which might later be violated. The high-level guidelines reported in Chapter 6 were combined with several new heuristics in order to generate and allocate a set of DDOs to two altered versions of *Andy's Garden*. Each altered version is meant to be played in a separate *games session* (i.e. a separate day of the evaluation, following a session in which the child has played the basic games).

Both the basic and altered games versions were tested with a small group of typically developing children, focusing on usability issues. These sessions identified a number of desirable changes to improve ease-of-use and to increase the interactivity of some parts of the games. Implementing these changes produced the final version of *Andy's Garden*.

7.2.2 Technical information

7.2.2.1 GameSalad game engine

Andy's Garden was developed in a basic (free) version of the GameSalad Creator game engine for Windows (GameSalad Inc., 2014). This is a drag-and-drop interface for game development, which largely uses the logic of object-oriented programming. GameSalad was selected because it includes many ready-made capabilities that would be important in this design, including a powerful physics engine, and ready-made options for creating, updating, and referencing tables of values. It can support animations in a limited way, if the user creates a separate image for every frame and sets them to display with a certain order and speed. GameSalad also provides an easy avenue for dissemination. The company maintains an online "arcade" where users can upload games, and others can play them for free in a standard web browser. This is a simple, accessible option to make the research software available to participants and the public. Further details of the GameSalad engine are given in Appendix G.

7.2.2.2 Graphics

The Andy virtual character is illustrated with 2-D images of a 3-D character model, "cartoon boy andy" (3deverart, 2010)[1], the same as was separately purchased for ECHOES. As the current games do not use 3-D rendering, *images* of the model were used rather than the model itself, with additional postures created by modifying the original images in Adobe Photoshop Elements (Adobe Systems Incorporated, 2010). All other background and object graphics were created by the author using the same program, with the exception of the cloud graphic (created for the ECHOES project, but further edited here).

7.2.2.3 Sound and character dialogue

Andy's dialogue was recorded specifically for this research, voiced by a child[2]. He was instructed to sound positive and encouraging, exaggerating key words. As Andy is physically static, voice is very important for shaping his perceived personality and role. Prior to recording, two colleagues with experience in autism and children's technology gave feedback on the planned dialogue, checking it for clarity and simplicity. Approximately 30 phrases were used in the final game. A full list of Andy's dialogue and information about non-voice sound effects is reported in Appendix G.

7.3 Initial design decisions and concepts

7.3.1 Design constraints and high-level decisions

Prior to generating any specific ideas, a number of high-level decisions were made that constrained the form of the eventual design, and some of its contents. The most important decisions are each addressed briefly below.

Scale of evaluation study with new games The current games were not intended as an intervention tool, but as part of a proof-of-concept study investigating child behaviour during game-play only. Far less content was needed than in a longer-term

[1]Purchased from http://www.turbosquid.com/.
[2]Many thanks to Leon Blake for playing Andy.

intervention study. Here, enough content was required to engage children for approximately 40-60 minutes of interaction. Based on ECHOES experiences about manageable session lengths, this was estimated as 3 sessions for most children. Thus, the amount of content created was intended to engage children for 15-20 minutes at a time, over multiple days.

Build on existing ECHOES activity concepts To best investigate the current questions of interest, it made sense to build on activities that had already been tested with a range of children and found to be likeable and of appropriate difficulty. There was considerable scope to borrow the features and premise of one or more ECHOES activities, and still produce something that interpreted them in quite a different way. There was also an opportunity to correct some things that were difficult in ECHOES, such as confusion between similar activities, and difficult object physics. A tentative goal was to develop at least one mini-game based on *Flower Growing (in flower pots)*, and an activity based on *Ball Sorting*. In ECHOES videos, these represented two different styles of child-system interaction (see 6.2.3.1) and were demonstrably popular with participants.

Use a garden context and related objects The choice of a Magic Garden setting for ECHOES was informed by PD with children and an adult stakeholder advisory group. Gardens were identified as being spaces that would be familiar, but were unlikely to have unpleasant or anxiety-inducing associations (as was the case with playgrounds, an initially suggested context). Children appeared to respond well to the Magic Garden in the ECHOES pilots and evaluations, and there seemed no reason to select a new context here.

As a related constraint, objects used in the main game play and in any DDOs were limited to objects that were broadly congruent with a garden setting. While incongruous objects might make equally effective—or even more effective—DDOs, this choice helped narrow the current number of design options. Garden-related objects were also thought to be reasonably familiar and reasonably neutral in association for the participating children.

Re-use Andy in a similar role to ECHOES The same Andy character model (see 7.2.2.2) was re-used in the new games, playing an ECHOES-like role. He was mainly present to help demonstrate activities, encourage the child, and make silly mistakes (via DDOs in altered game versions). His ability to respond to the child—other than scripted feedback—was minimal. It was anticipated that the adult would be the main social partner with respect to responsiveness and concrete assistance. Playful behaviours from Andy (i.e. in the style of the *Tickle Andy* activity in the original ECHOES, see A), such as responsiveness to touch or objects, were identified as desirable additions to his behaviours, if possible.

Activities will not teach skills The ECHOES activities tried to teach, or at very least promote practice of, social skills such as turn-taking and gaze-following. The new games did not have skills-teaching or within-game skills practice. The only interest was in the communication and social behaviours that children spontaneously produced within and around the technology.

Turn-taking will be a possibility, but not an explicit goal None of the new activities will have turn-taking as an explicit goal, but should keep the possibility of Andy taking activity turns. Aside from modelling activity actions, this is a good opportunity for Andy to make mistakes (previously successful at motivating initiations). However, turn-taking or coordinated action should not be required to meet activity goals, as it is not a particular interest of the current work.

Create a simpler, more scripted environment In the original ECHOES activities, some of Andy's behaviours which were reactively planned could feasibly have been scripted and still resulted in a similar experience. Here, the decision to script Andy and environment behaviours was taken in the interest of simplicity, and because it was not expected to cause a noticeable loss of functionality. It was also far easier to control children's exposure to DDOs in a pre-planned environment.

Andy's actions also have become simpler because the new environment was 2-D, using images of Andy rather than a 3-D rendering. Animations of his movement were minimised, and he was fairly stationary. This meant that"new Andy" compensates with more magical actions to make up for not moving around (e.g. see Section 7.4.6).

Child and adult stakeholder involvement The current work sought relatively limited involvement from children with ASC, in *user* and *tester* roles only (Druin, 2002). The current project built directly on ECHOES and retained the same general context, VC, pace, and style of interaction—all of which were already developed through extensive, iterative PD with children and stakeholders (Frauenberger et al., 2011, 2012a, 2013; Porayska-Pomsta et al., 2012). Design participation of children with ASC (and stakeholders) is inherently a “scarce resource”, both at the level of high researcher demand for a relatively small pool of people, and in terms of what may be practically and emotionally manageable for individuals. The core characteristics of ASC clash with core elements of design participation in many respects (as discussed in more detail in Frauenberger et al. (2013)). For this particular project, conducting additional PD beyond that in the original ECHOES was not a conscientious use of resources and would have added limited value.

7.3.2 Growing game concepts

The design for this game is based closely on *Flower Growing (in flower pots)* in the original ECHOES. As discussed in 6.2.3.1, both games that used the magic cloud to grow flowers allowed a child to hold down the cloud and drag it around, making rain and growing flowers for a long period of time. This often led to a flow-like state, from which children were not easily distracted to interact with a partner. In short, there was a very different activity pattern to the small, repeated “units” in *Ball Sorting*, that provided breaks for the child or researcher to initiate communication, as well as mini-achievements when each unit was completed. The current design planned to follow the original *Flower Growing* closely, but to try to introduce effects and rewards that might break up the activity and provide additional feedback—an attempt to split the difference between long, continuous actions and discrete units of action.

7.3.3 Sorting-based game concepts

The design for this game starts from *Sorting balls into boxes based on colour (Ball Sorting)* in the original ECHOES. This was a very popular game, with many children requesting it session after session. It provided good opportunities for social interaction due to its units of repeated action, had attractive sensory rewards, and was home to almost all of Andy’s most noticeable and amusing mistakes. Children immediately understood its premise, often without any need for instruction or demonstration. Unfortunately, it also generated many interactions about difficulties with the touch-screen and game physics (e.g. children repeatedly dropping objects, objects ricocheting off one another). Creating a closely related game was thought very likely to interest and engage children in the same age and ability group, and to be at the right level of com-

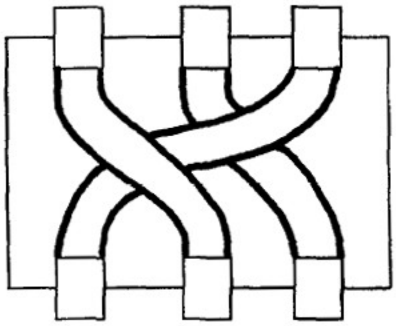


Figure 7.1: The tubes task apparatus, reproduced from (Hood, 1995, p.581).

plexity.

A new design based on *Ball Sorting* provides many options for which objects might be sorted, along which dimensions. Literature on the "tubes task" (Hood, 1995, 1998; Russell et al., 1999) suggested a new potential structure for the sorting, using a framework of crossing, opaque tubes rather than boxes (see Figure 7.1). The tubes task is a developmental psychology method (or rather, a physical apparatus) used to study young children's emerging concepts of gravity. Balls are dropped into the tubes, and researchers study children's errors in where they look for the ball at the bottom (Hood, 1995, 1998; Russell et al., 1999). The current work is not interested in gravity; what is relevant is that this task shares many features with the original ball-sorting task (e.g. repeated units of activity, cause-effect relationships) but the period in which the ball is hidden in the tube suggests additional opportunities for establishing and violating child expectations

Two game ideas were developed around the premise of sorting objects into tubes: sorting apples[3] by colour, or by size. Both ideas took as their setting a garden with an apple tree, and a tubes apparatus (similar to Figure 7.1) with baskets or containers at the bottom. As in ECHOES, Andy would announce the activity, identify its goal with a short utterance (e.g. "Will you help me sort the apples?"), and demonstrate the main unit of action by taking a turn. As he needs to be stationary on screen, he might sort an apple by throwing it, or using some kind of magic. The child would then continue these units of action, perhaps with Andy taking additional turns. There could be sound or visual effects when the apples go in the tubes, are sliding down tubes, and appear in the baskets below. Once sorted via the tubes, apples are inaccessible and cannot be

[3]Apples were chosen as the sortable item because they are roughly round and ball-like, are familiar to children, naturally occur in different colours, and are congruent with a garden setting.

removed from baskets. When all apples are sorted, there is a sensory reward of some kind, and the tubes arrangement disappears, leaving the baskets. Andy praises the child and thanks him for helping.

Early design concepts explored two sorting dimensions: colour, and size. If sorting by colour, there would be two apple colours (red and green) but three tubes. Apples could not go into tubes of the wrong colour, perhaps just sitting on top, or, bouncing back to ground. Both colours could go into a magic third tube (shown as white in Figure 7.2) and be transformed in some way (e.g. switch apple's colour). If sorting by size, the apples would be large and small versions of the same colour. There are only two tubes, each a different size, and now "transparent" so children can see what happens inside. The game play would be the same, except that now putting apples in the wrong tubes could create an opportunity for the child and Andy to solve a problem, for example if a big apple gets stuck in the small tube (illustrated in Figure 7.3).

Both the idea of a transformation-making tube and items becoming stuck in a tube (and requiring a child-character cooperation to remove) were generated during a short design workshop conducted with six colleagues experienced in human-computer interaction or ASC-tech. They discussed the tubes task (mentioned above) as the potential basis for child-virtual character interaction, and suggested that things getting *stuck* in tubes could create a useful problem-solving or helping opportunity. Aside from the specific suggestions above, the workshop provided no further input into the eventual designs reported in this chapter—partly because the research questions evolved from those projected at the time of the workshop. Consequently, it is not discussed further.

Informal critique of these two sorting ideas with ASC-tech colleagues identified that developmentally young children might struggle with relative size perception (e.g. larger and smaller). There is some related evidence to support this concern (e.g. May and Macpherson, 1971). Thus, they might find it difficult to sort objects on this dimension, especially if the objects were otherwise the same. This could seriously hinder some participants' ability to do the activity, and establish expectations about its rules. In light of this concern, the colour-sorting concept was chosen for final development.

7.4 Basic versions of mini-games

This section gives the main description and visual illustration of the *Andy's Garden* game suite. It focuses on presenting the final versions of each mini-game, except where there were particularly important points of change or evolution. On the whole,

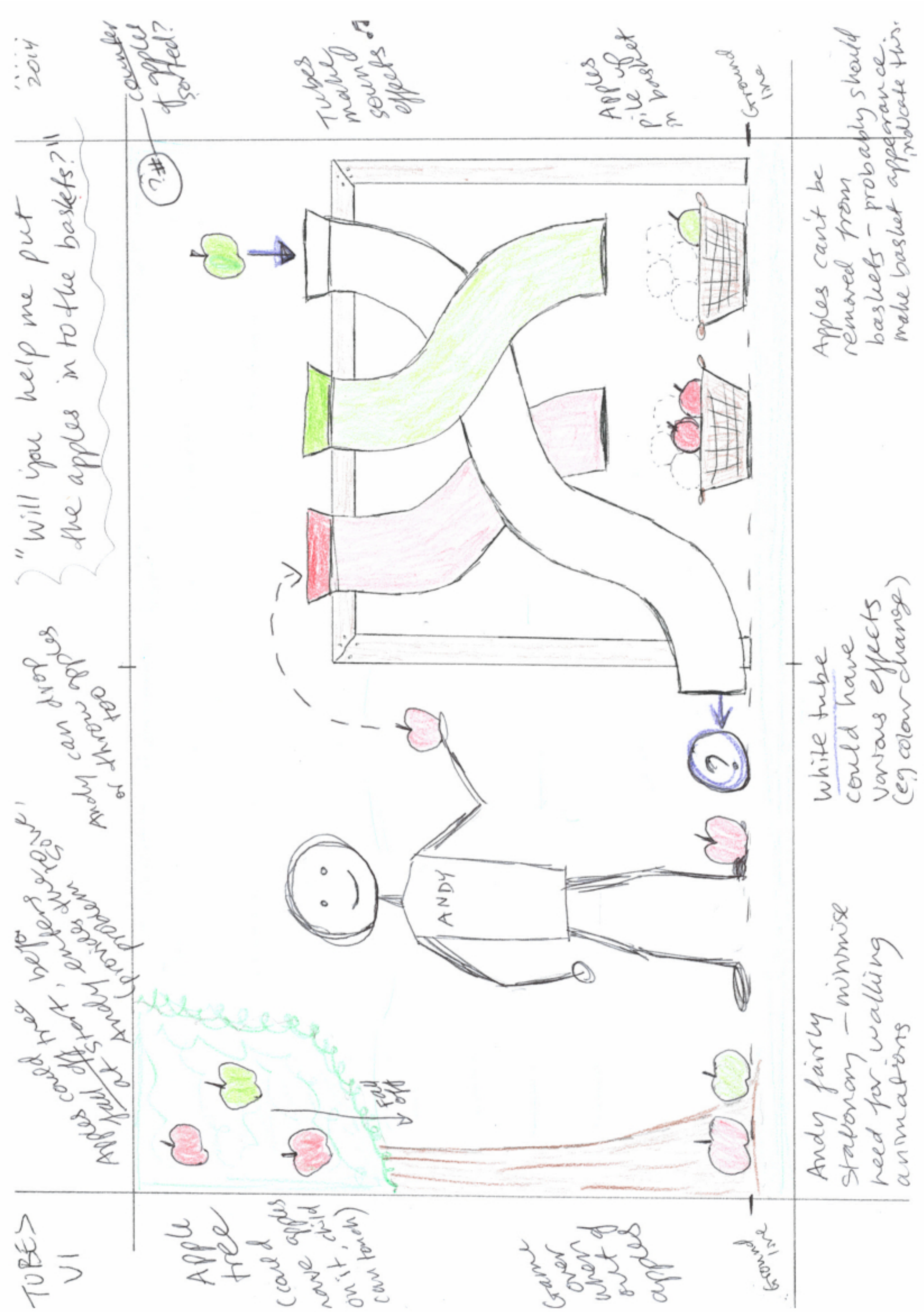


Figure 7.2: Concept sketch for an ECHOES-like game based on sorting apples by colour.

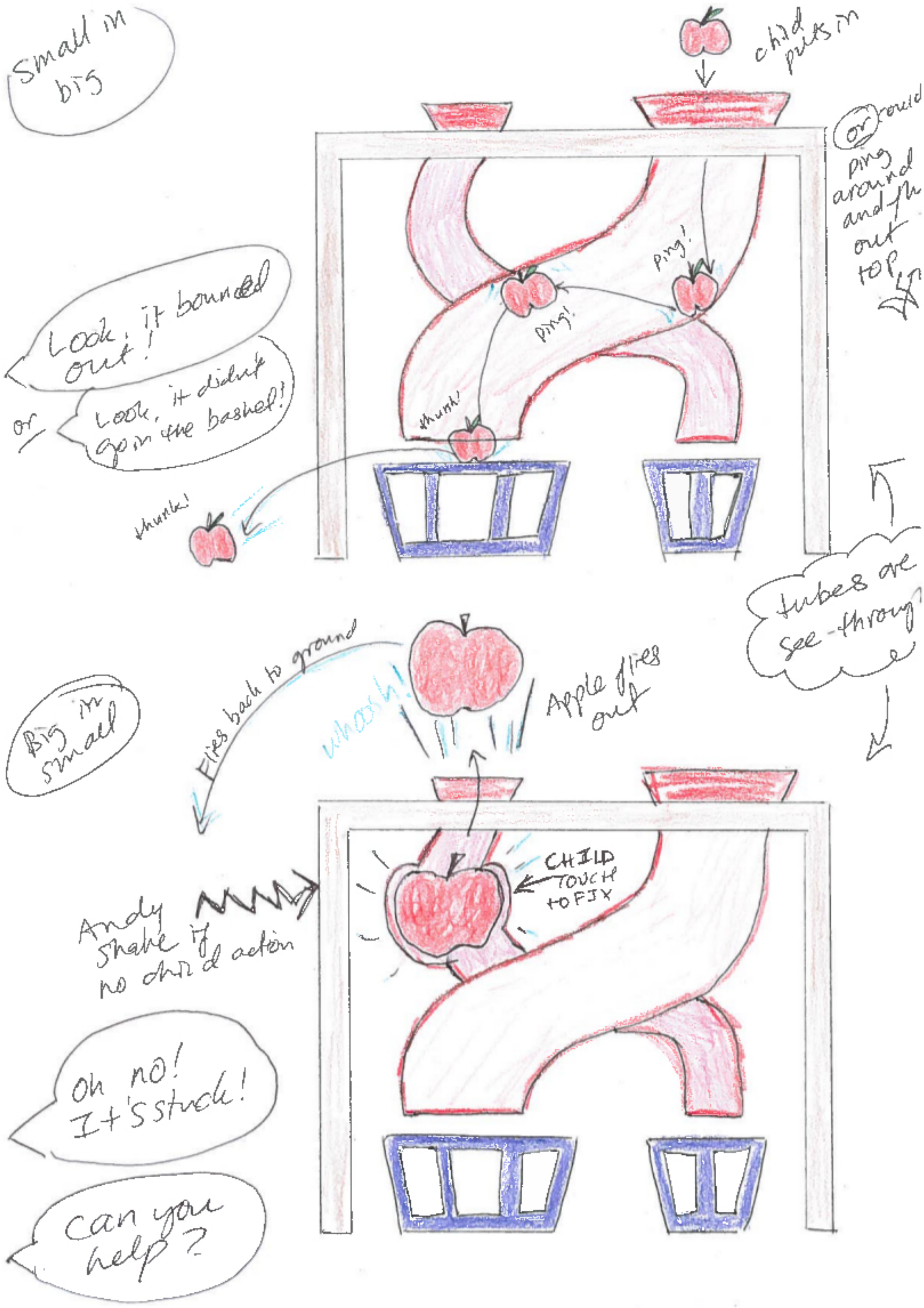


Figure 7.3: Detail sketches for what might go wrong when apples are sorted into tubes of the incorrect size.

the mini-games follow fairly directly from their original concepts. Each activity is described chronologically, from its opening material, through the main game play, and the concluding material. Special issues with the design and child-system interaction are described separately.

At several points, specific elements or decisions are identified as having come from design critique input. This critique took place about two thirds of the way through developing the basic games versions. Here, each mini-game had its final graphics and basic mechanics, but was missing a number of Andy's behaviours and interactions. These incomplete versions were discussed in a design critique with academic researchers[4]. Specific outcomes and issues from this are reported in the description of the basic games versions where they are most relevant, rather than discussed in isolation. The following section gives further details of the expert critique, prior to the game descriptions.

7.4.1 Design critique with HCI and children's technology experts

In the expert design critique, six researchers with expertise in human-computer interaction, children's technologies, and autism participated in a 45-minute session, as a part of a research group meeting[5]. After a brief introduction to the planned use of the games as a research tool, individuals or pairs played through the three mini-games. They were given the following prompts about which to discuss and make written comments:

- "Comments about overall look-and-feel, pacing, overall interaction"
- "Comments about Andy and his behaviours" and a suggestion to try waiting and not doing anything for a while, to trigger Andy's turn-taking
- "Comments about objects, object effects, player interactions with objects"
- "Any suggestions for things to add, subtract, or change?"

This free-play and discussion part of the critique provided positive feedback on the state of the games at that point; the design was thought to be substantially appropriate for its intended purpose and the target group of children. Critique participants identified a number of desirable changes to make the games more easily playable, and more understandable to young children.

[4]Henceforth referred to as the expert critique.

[5]This was a completely separate event to the design workshop mentioned earlier in the chapter.

There were also two targeted discussions on tricky design issues. First, how could Andy demonstrate activity actions so that children would understand what *they* needed to do, while he remained fairly stationary on-screen? This is addressed in section 7.4.5. The second discussion explored Andy's dialogue, and how to succinctly introduce game goals and give instructions. The games description returns to this issue in several places.

The critique also highlighted an important logistical issue: playing the games over wi-fi could be extremely slow, on some devices unplayably slow. Changed event timings disrupted the causal relationships between actions and results. For a reliable in-school evaluation study, the game would need to be run directly off the evaluation computer in GameSalad's preview mode, rather than via the online GameSalad arcade.

7.4.2 Games sessions and order of play

Andy's Garden consists of three mini-games, a menu screen, and a hello/goodbye scene with the VC Andy. When a child plays through a games session, s/he will see them in the following order, with a black screen also appearing briefly in between to help signal the transitions. The order of play was the same for basic game versions (session 1), and altered games versions (sessions 2 and 3).

1. Menu 1 (flower icon active)
2. Andy welcome sequence
3. Flower Growing[6]
4. Menu 2 (apple icon active)
5. Tidying up the apples
6. Menu 3 (carrot icon active)
7. Carrot Growing
8. Andy goodbye sequence
9. Further play: must exit to game editor and re-launch. Dynamic menus are in effect and Andy welcome sequence will play again each time *Andy's Garden* is launched.

[6]Due to their similarity, it was an arbitrary choice to present flowers first rather than carrots.

Each scene is described below, in terms of its appearance, goals, and possible actions. These descriptions all refer to the basic game versions: those which have "baseline" behaviours and are used to establish child expectations. While these game versions do not include DDOs, they do include many *inherently novel* aspects, because they are the child's first encounter with Andy and the rest of the game content. Thus, there are opportunities to detect novelty. Altered games versions, with DDOs, are described later in the chapter in Section 7.5.

7.4.3 Menus

The menu has no text, only a large icon for each of the three games (flower, apple, and carrot). The menu is deliberately related to the *visual schedules* that many children with ASC use in their classrooms or home, to help them understand and anticipate transitions (e.g. Dettmer et al., 2000). Commonly, a series of small pictures indicate upcoming activities for a day, a child's chores, or similar. These may be displayed at the front of a classroom, representing the day's schedule, or may be in an individual child's work space. Removing pictures of "finished" activities and reviewing what is coming next can help to aid transitions. The games menu works the same way, with ordered images representing what is happening now, next, and later. It highlights the (active) icons for the current activity, and showing disabled, greyed-out icons for the other activities (illustrated in Figure 7.4). At any one time, only the button for the next activity is presented in colour, and will produce a result. The game returns to the menu between each activity, so the child can see what is coming next, and move on when s/he is ready. It is not possible to re-start *Andy's Garden* or re-play activities via the on screen menu. The researcher must exit the GameSalad preview mode and open the game again.

The menu for *Andy's Garden* has no equivalent in the original ECHOES. There, sessions did not have a pre-set activity order, and the researcher used a GUI to select activities. Here, there are limited researcher controls via keyboard hot keys (described in Section 7.4.7). Aside from its benefits as a simple visual schedule, placing the menu on the main screen and allowing the child to start each activity (even if the order is set) gives the child more ownership and control in the interaction.

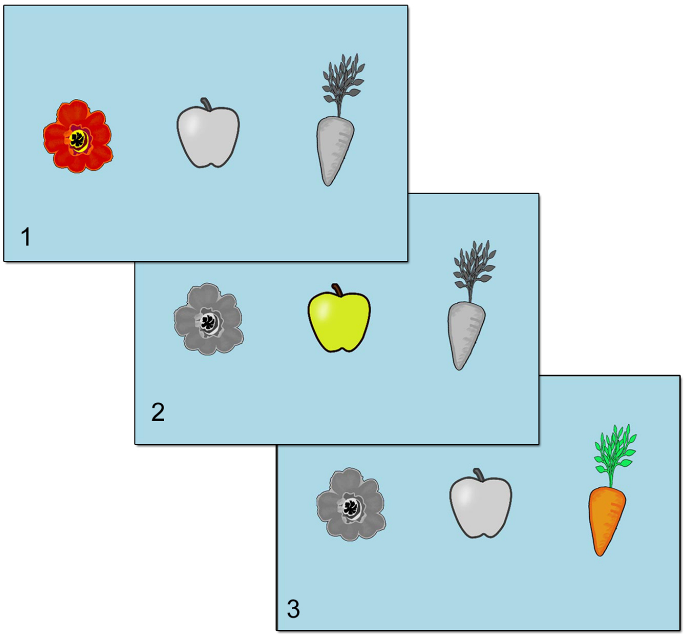


Figure 7.4: Menus for *Andy's Garden*, as they display before the flower game (1), before apples (2), and before carrots (3).

7.4.4 Opening and closing scenes

In both of these scenes, Andy stands alone on a grassy background, approximately centre screen. No other objects are present, in the hope of directing the child's attention to Andy immediately (see Figure 7.5). In the opening scene, he waves after a few seconds, and says *"Hi, I'm Andy. Will you play with me?"*. After a pause for the child to process this information and/or respond to it, he continues *"Let's play in the garden."* After another few seconds, the scene transitions. In the goodbye scene, Andy speaks before waving, saying *"I need to go home now. Let's play again soon. Bye!"*.

The opening scene with Andy will be the first thing to which the child is introduced in the environment, after the menu. It is repeated at the start of every session—even more often if children play through the games multiple times in one session. The opening and closing scenes never change (i.e. have no DDOs in later games sessions). Their unchanging repetition is intended to be part of the sameness around the main game content. Additionally, for children in this age group who can be quite excitable about leaving the classroom to play games, the opening provides another small period of leeway for children to finish talking, settle down, and and attend to the screen.

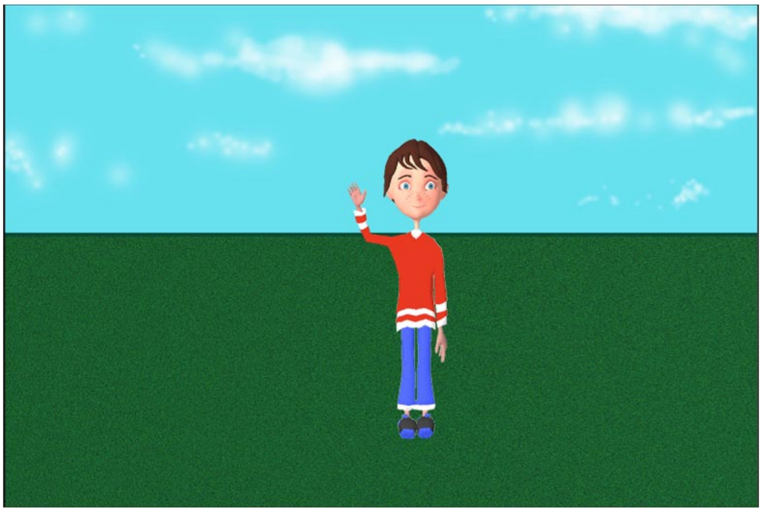


Figure 7.5: The *Andy's Garden* hello and goodbye scene.

7.4.5 Flower Growing and Carrot Growing

Both of these games implement the initial growing game concept. After test implementation of key concepts in flower growing, it was decided to increase the amount of content available for the evaluation (and minimise total development complexity) by implementing two versions of a growing game, with the same basic mechanics and slightly different action options. Elements and actions specific to only one game are noted as such.

Like *Cloud Raining* and *Flower Growing* in ECHOES, the child *can* engage in long continuous actions with the cloud, going from one plant to the next without stopping. However, as noted with the growing game concept, this design tries to introduce markers of progress and potential stopping points into an otherwise continuous activity. The special effects around finished plants try to unitize the activity more, splitting the difference between the completely continuous *Cloud Raining* and the discontinuous, incremental sorting activities. The hope is to draw on the engagement benefits of one, and the interaction opportunities of the other.

Game opening Andy begins center screen, standing between two planter boxes, with the magic cloud on the left side (Figure 7.6, top). In ECHOES, children or the researcher often talked over the first seconds of a new activity, losing Andy's entrance, utterances or action demonstration. Here, Andy does not do anything immediately—there is a deliberate pause in which children can orient themselves to the environment and finish any communications that were happening about the previous content.

After this pause, Andy says "Look, a cloud! Let's grow some flowers" or "...Let's grow some carrots". While speaking, he looks up at the cloud, then back out to the front (apparently at the child) and smiles. If the child is not already touching the screen, Andy takes a turn to demonstrate how the cloud can make rain and grow the plants. He says "My turn!" and the cloud moves (magically) to hover over one of the six plant areas (chosen randomly). In flower growing, these are bare earth. In carrots, they are each marked by a tiny sprout at the start of the game (see Figure 7.10).

When Andy takes his turn, he looks up right or up left to look at the cloud, and extends his arm to touch it. On his touch, it lights up with a blue outline and a rain sound effect plays (visual and aural feedback). This additional, visual feedback was suggested in the expert critique for the child's benefit, so that she knows her action is correct. When active, the cloud produces a rain animation (Figure 8.9). A plant sprouts where rain touches the dirt in the planters. Andy then retracts his arm, looks forward again, and says "Wow!". He waits for the child to process what has happened, and do something. Figure 7.6, bottom, gives an example of Andy's turn taking midway through the activity, when several plants have been started. If a child begins using the cloud immediately when an activity opens (without waiting for his demo), Andy will skip his opening dialogue and go straight to giving feedback.

Andy's turn taking: How to model child actions? The early decision to make Andy fairly stationary while on-screen ran into a conflict with the need for him to demonstrate the cloud. He needed to be able to reach a cloud that might be all the way at the edge of the screen, and *also* demonstrate in a way that would convey to children what *they* should do. The apparent solution seemed to be for Andy to extend his reach in some way, to touch the cloud. Using a garden-related or rain-related object (e.g. rake, umbrella) would solve one problem, but suggest *a completely different task* to a watching child. The first goal would appear to be getting the special object, and *then* using it to take a turn. If the object was not accessible to the child, this could set up an immediate problem of re-directing the child to the cloud.

If objects were ruled out, this left two more possibilities: some kind of magic effect between Andy's hand and the cloud, or actually extending Andy's arm, cartoon-fashion. These two options were presented and discussed during the expert design critique, as illustrated in Figure 7.8. Per the critique discussion, neither option was seen as ideal. Some colleagues found the arm extension rather off-putting, but agreed that it would succeed as a demonstration, and was not dissimilar to what children

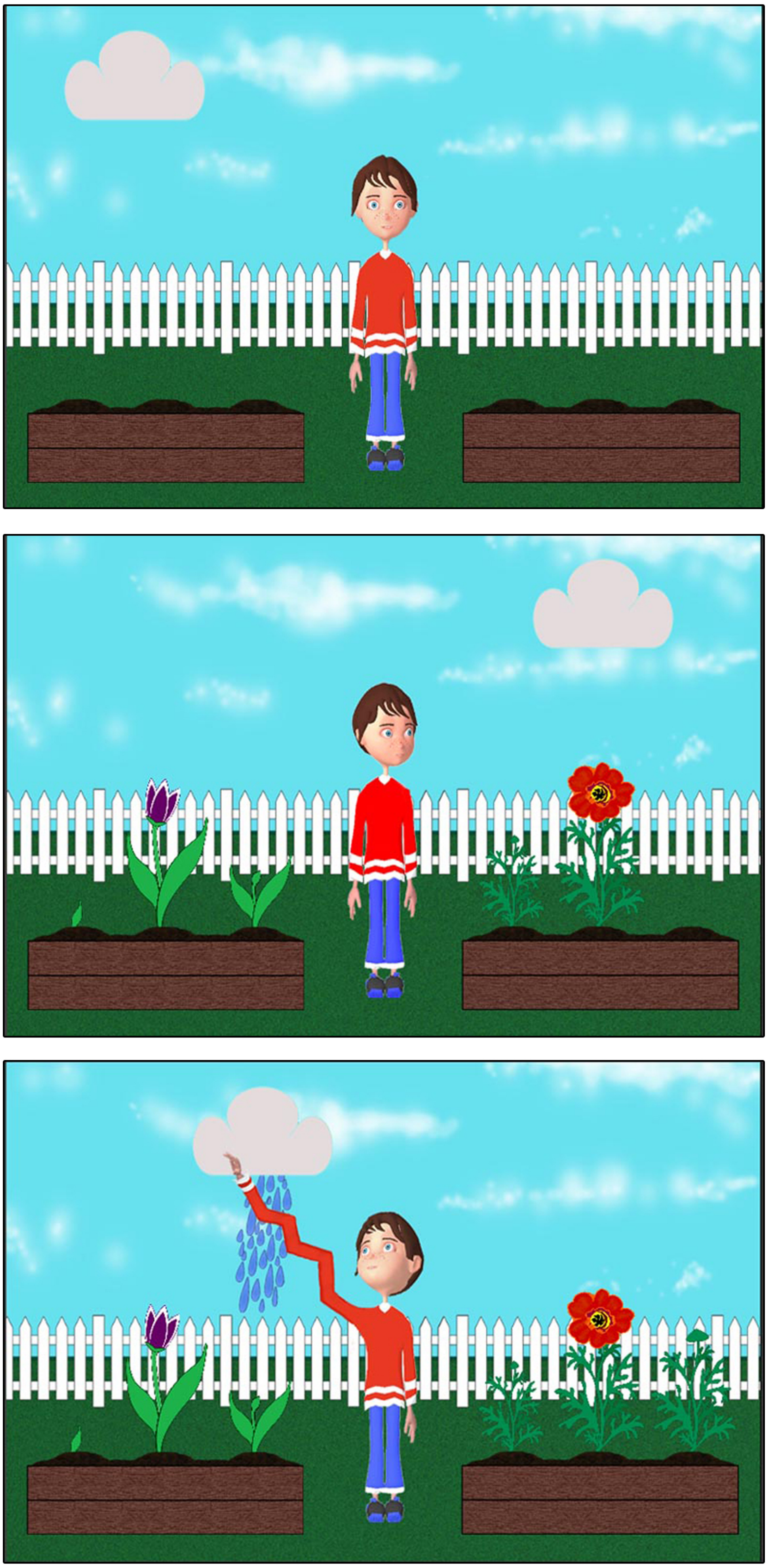


Figure 7.6: *Top:* The child's opening view of the flower game, prior to Andy demonstrating the cloud. *Center:* Andy turns to look at a red flower the child has just completed, saying "It's a red flower!". *Bottom:* Andy takes a turn partway through game play, extending his arm to reach the cloud.

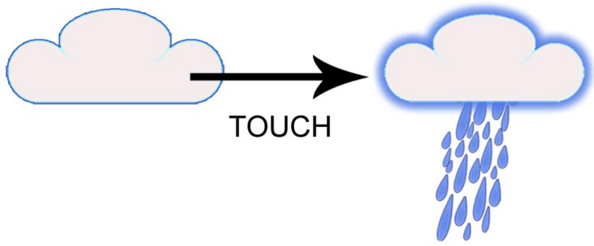


Figure 7.7: The cloud provides visual feedback when touched, in addition to sound feedback

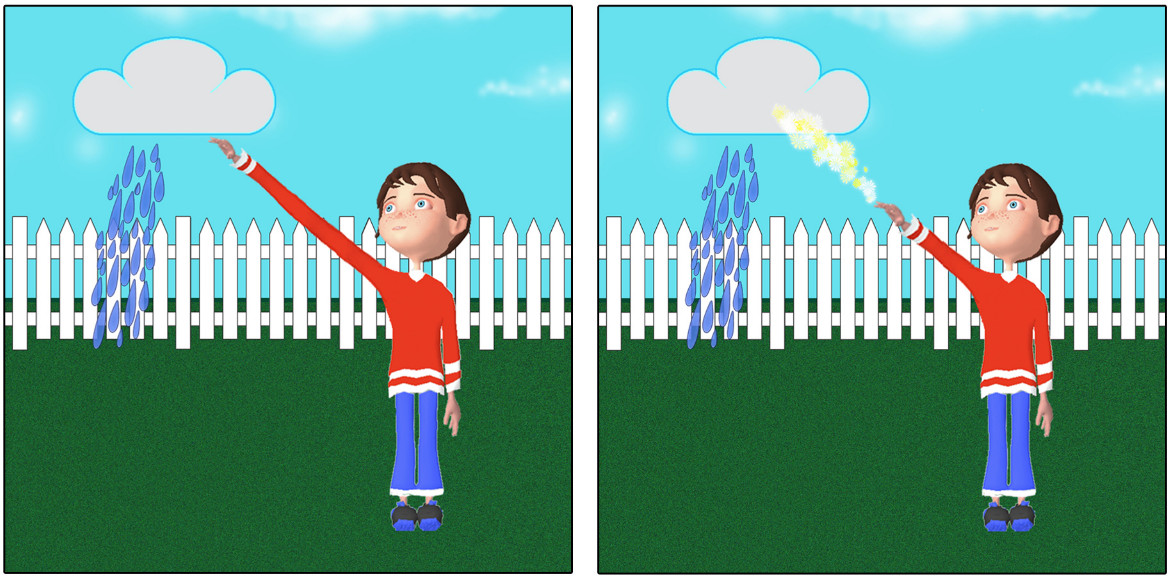


Figure 7.8: Two options for how Andy could demonstrate cloud use, presented as a basis for discussion in the expert critique.

might see in animated media. Andy actually touching the cloud *with his hand* was deemed important for conveying what the child player needed to do, so the cartoon arm extension was eventually chosen and implemented. He extends his arm in stages, accordion-like, as in Figure 7.9. This is shown in situ in Figure 7.6, bottom.

Main play Making rain with the cloud is the main action available to the child, keeping the activity very focused. When the child touches the cloud, the rain is generated in "units" rather than as a stream. Each unit of rain increases a plant's growth by one stage. There are six stages for each plant (see example in Figure 7.11). This may sound like a large task when multiplied by six plants, but the cloud only needs to be

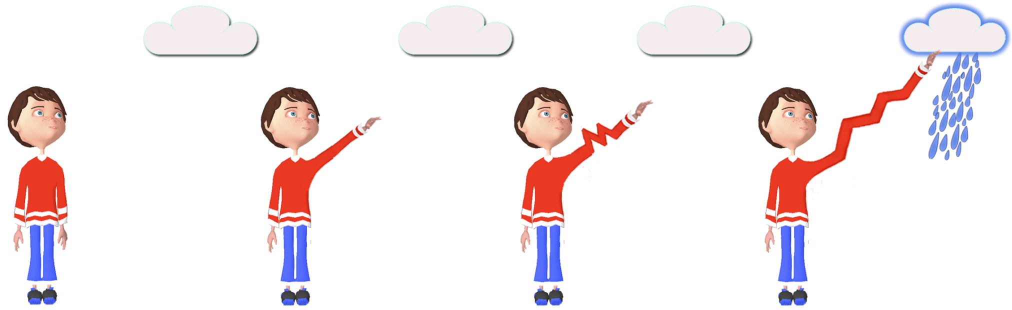


Figure 7.9: Stages of Andy's arm extension to take a turn using the cloud.

continuously held over each plant for about 7 seconds to grow it completely. Even with some starting and stopping, the whole activity takes only a few minutes. Andy is programmed to give positive feedback every few seconds while the child is using the cloud, saying "Wow!" "Good job!" and "Well done" (chosen randomly).

When the child finishes a flower, its bud blooms and a fanfare sound plays. Andy will turn to the relevant screen side (appearing to look at the flower) and then back, smiling and commenting on what the child did (e.g. "It's a [colour] flower!" or "Mmm, a carrot!"). Figure 7.6 (centre) gives an example. When carrots are finished, they light up with an orange outline, similar to the visual feedback on the cloud (see Figure 7.10, right). This feedback was suggested in the expert critique, as participants pointed out that (unlike flowers) it was unclear when carrots were finished growing.

The finished plants each have special properties. If the child touches a carrot, it flies up out of the ground with a satisfying *pop* (as shown in in Figure 7.10, right) and falls back to bounce on top of the planter. Should the child move the carrots around the screen, s/he might happen to touch the cloud. Touching the cloud with a carrot makes it orange; touching it again returns it to white (illustrated in Figure 7.12). Later, picked carrots were given additional actions at the request of child testers (see Section 7.6). Flowers cannot be picked, but if the child presses on them they "dance" by wiggling from side to side.

In order to grow all the plants, the child must at some point pull the cloud across the screen from one planter to the other, going over Andy. If Andy is rained on, he reacts by exclaiming "it's raining on me!". If the child keeps raining on him, he will keep saying this up to a maximum number of times. This cut-off point is to eventually re-direct the child back to the rest of the activity, should s/he find it especially rewarding to rain on Andy. Andy gained some additional actions as a result of child tester feedback,

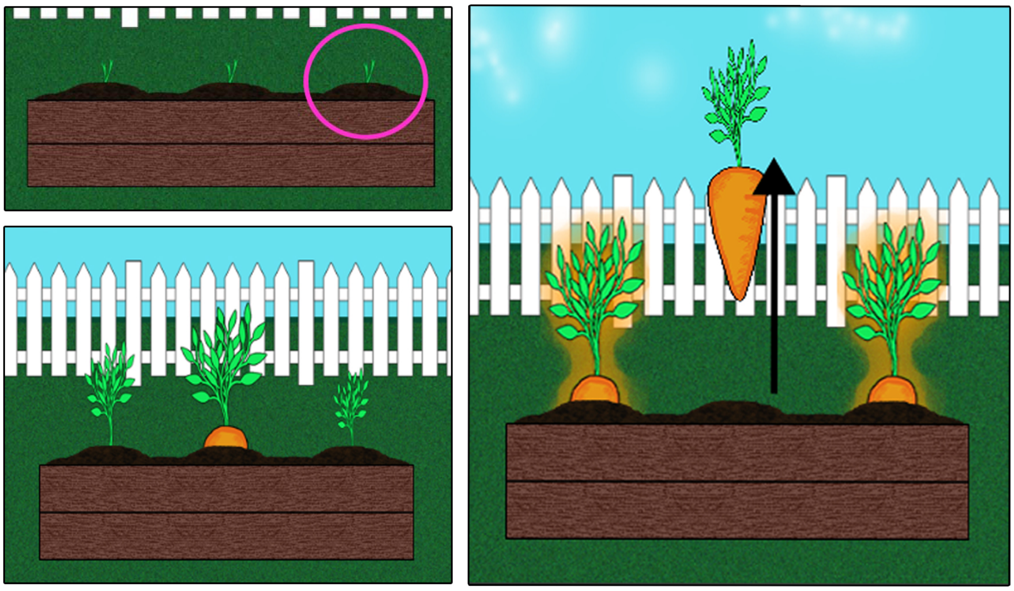


Figure 7.10: Details of *Carrot Growing*. *Top left:* Sprouts present at the start of the activity. *Bottom left:* Carrots in various stages of growth. *Right:* Completed carrots light up with an orange outline. They pop out of the ground when picked via a child's touch.

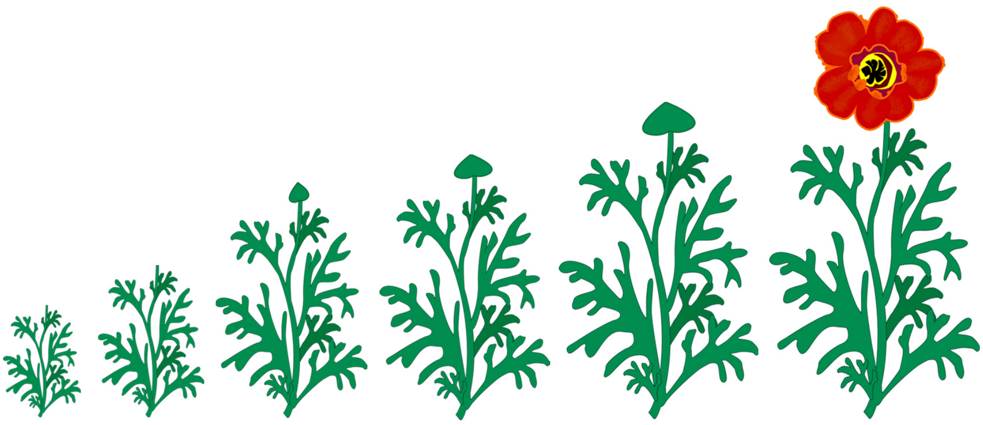


Figure 7.11: Example of the six stages of growth for the red poppy in *Flower growing*, with the flower blooming in the final stage to indicate it is completed. All flowers and carrots have six stages.

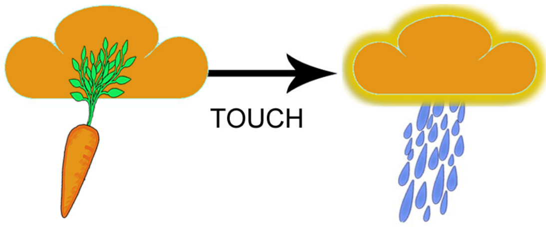


Figure 7.12: When a carrot touches the cloud, it turns orange. If the cloud is still active (i.e. not all carrots complete) the cloud has an orange outline when used to make rain. Touching the cloud with a carrot again will toggle it back to white.

described in 7.6.

Game conclusion When all plants are completed, there is a short pause before the child gets a reward animation: bright, animated bursts of stars appear over each planter box, accompanied by the sound of fireworks and people cheering (Figure 7.13). These last for several seconds and then fade away. This reward was inspired by the fireworks reward in the original *Ball Sorting*, which had been particularly popular. When the reward begins, the cloud is disabled and cannot be used to make more rain.

After the reward, Andy smiles and says “Finished”. Following another short pause he re-states what the child has done (“You grew all the flowers!”), bookending his goal announcement from the start of the activity. After another short pause, there is a black cut scene and the menu re-appears.

7.4.6 Tidying up the apples

This game implements the original concept of a sorting game using tubes. During the expert critique, a discussion of Andy’s dialogue identified that “sorting” might not be a familiar word for young children, even if they were perfectly familiar with the *idea* of sorting things in class or at home[7]. Several expert critique participants advised that the activity might be better introduced in terms of “tidying things up” or “putting things away”. Thus, this mini-game became “Tidying up the apples” and was described in this way to children, during the evaluation study[8].

[7]This agrees with what happened in Study 1: researchers referred to the *Ball Sorting* activity, but children did not. They tended to refer to this activity by the objects involved (boxes, balls) even if the researcher *said* sorting.

[8]In the rest of this document, it is still frequently referred to as apple sorting, or the apple game.

Figure 7.13: The end-of-activity reward as it appears in the flower and carrot games. Several seconds of animated starbursts are accompanied by the sound of fireworks and people cheering. The reward appears almost immediately after the final plant is completed, and before Andy says his end-of-activity dialogue.

Game opening Andy begins left screen, next to the apple tree. As the activity loads, six apples fall off the tree and bounce across the ground (Figure 7.14). After a pause, Andy smiles and says “Let’s tidy up the apples. Put an apple in the tube!”. Several seconds later, he smiles again and says “My turn”. Andy uses magic to make a (randomly selected) apple disappear from the ground in a flash of light, and reappear in his now-outstretched hand.

Andy’s turn taking As with the growing activities, Andy needs to demonstrate apple-sorting from a stationary position, and without using any additional objects that might confuse his demonstration for the child. In the early game implementation, Andy crouched down to the right or left, and picked up an apple (Figure 7.15). He then stood up and threw it to the correct tube. This was a clear demonstration of the action, but its implementation was very fragile and could be easily disrupted by any other game actions taking place during Andy’s turn. In the absence of a robust implementation for this action sequence, the first part of the sequence was changed. Now, Andy remains standing, and “summons” an apple into his hand. It vanishes in a flash of magic, and reappears in his hand with another flash. He then throws it to the tubes. This sequence is illustrated in Figure 7.16. In the altered game versions, several DDOs cause Andy

Figure 7.14: At the start of *Tidying up the apples*, the apples fall off the tree and bounce across the ground. One of the altered game versions leaves some apples on the tree as a DDO.

to throw the apple to the wrong colour tube, or throw and miss the tubes entirely.

Main play Children can sort the apples, and also explore their physical properties by tossing, bouncing, and stacking them. Apples will not go into the wrong colour tube; they simply fall to the ground and bounce. When an apple is dropped into the matching tube, it disappears and there are several seconds of a silly sound effect to suggest the apple zooming down the tube. Then, an apple re-appears in the basket at the bottom with a fanfare (the same as for a completed flower or carrot). No apple can be removed once in the basket. After each sorted apple, Andy turns his head toward the tubes and then forward (i.e. looking "at" child) and says a randomly selected feedback phrase[9] (the same as reported for growing games). If there is a long delay in which the child does not successfully sort any apples, Andy may take another turn. The child cannot give Andy any apples as a social overture or cue to take his turn; there was no easy way to implement this.

Both apple colours can go into the white "magic tube" (see Figure 7.17, top). This tube makes a tinkly, xylophone-like sound. After a few seconds, it outputs an apple of the opposite colour. If the child wishes, s/he can change all six apples to be the

[9] Unless multiple apples are picked up and sorted together, in which case he only gives one piece of feedback.

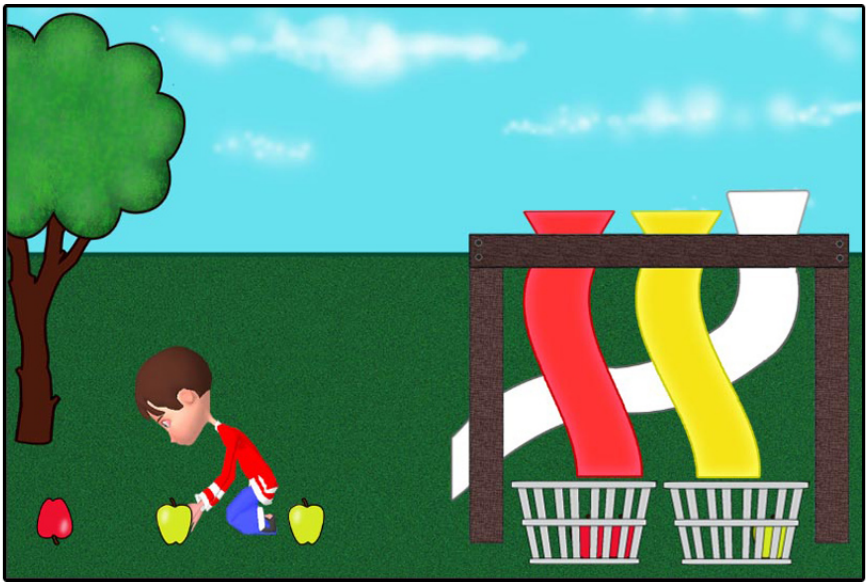


Figure 7.15: Early versions of apple sorting had Andy take turns by bending down to pick up an apple, then throwing it to the tubes.

same colour and sort them into the same basket. Apples can be changed an unlimited number of times.

The expert critique identified that in the original tubes configuration (Figure 7.17, bottom), the tops were close together and it was very easy to accidentally put a yellow apple in the white tube, or vice versa. Apples disappear instantly when they touch the top part of any permitted tube—there is no delay, and barely touching the top piece is enough. Especially as young children may have imprecise motor control, this was flagged as a likely source of confusion and frustration, if apples disappeared into the "wrong" (not intended) tube. The tubes were changed so that their tops were further apart, and also distinguished by height. Now, in order to use the white tube, the child must move an apple all the way up to it and cannot touch it by accident (Figure 7.17, top).

Game conclusion When the last apple is sorted, there is a brief pause and then the reward animation appears over the tubes structure. The animation is the same as in the growing games. When the stars fade away, the tubes also disappear. Andy finishes the same way as in the growing activities, except that he now congratulates the child "Well done, you tidied up all the apples!".

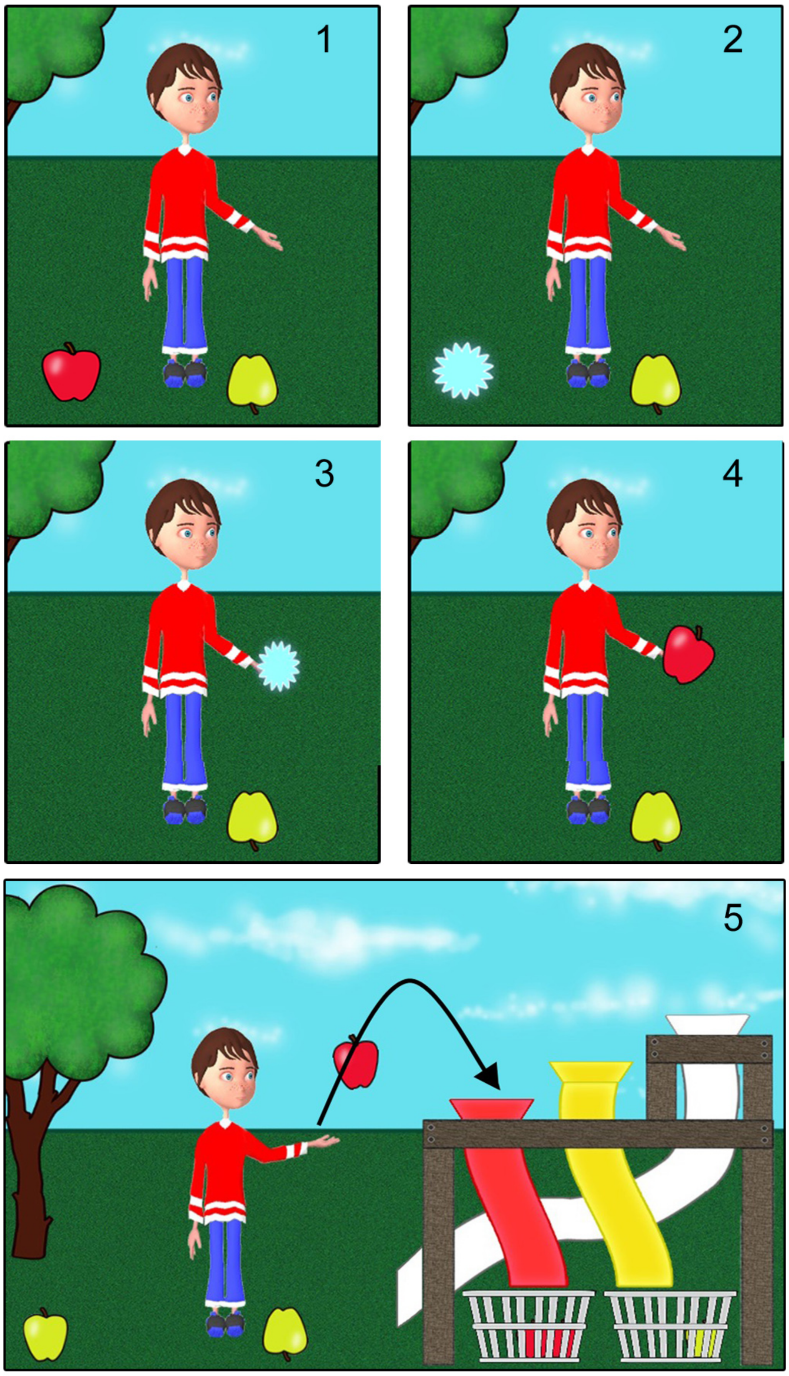


Figure 7.16: In the final versions of apple sorting, Andy "summons" an apple into his hand with a flash of magic.

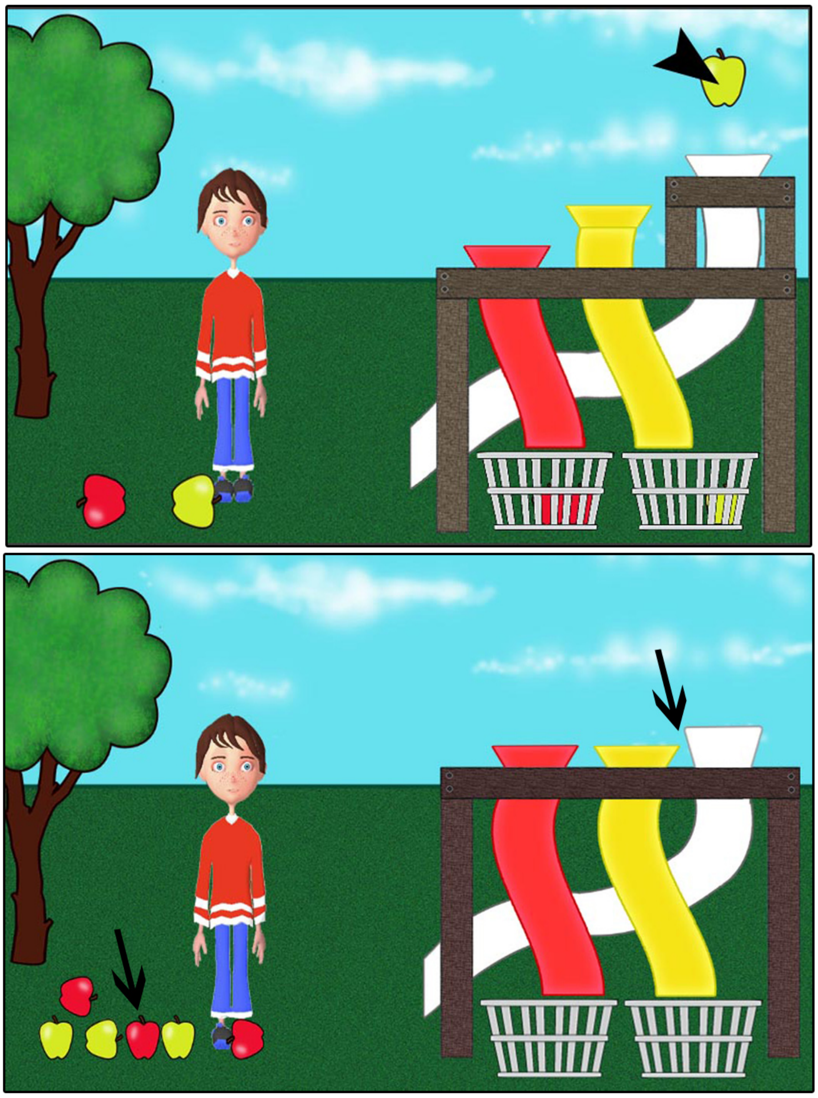


Figure 7.17: *Top:* The final version of tidying up the apples, with cursor indicating where the child would be holding an apple, about to drop it into the white tube. Here, several apples have already been sorted, and appear in the baskets. *Bottom:* Early, expert critique version of apple sorting. Black arrows indicate two key changes. Apples became larger, for ease of manipulation on the touch-screen, and the tubes apparatus was altered to create more space between the tops of the tubes (harder to make mistakes).

7.4.7 Researcher controls

Andy's Garden replaces the researcher GUI present in ECHOES with a set of keyboard hot keys. Both the original GUI and the hot keys allow the researcher to trigger limited character actions on demand (e.g. Andy take a turn). This gives a small amount of control over the environment, but is in no way a Wizard-of-Oz type control. Advancement between activities, which was a main function of the ECHOES GUI, is transferred to the child in *Andy's Garden*, via the touch-screen menu. The keyboard hot keys, listed in Table 7.1, allow for some additional responsiveness if the child tries to communicate with Andy (e.g. make him say "Thank you"). More frequently, the hot keys allow the researcher to intervene without *appearing* to do so. With a key, Andy can prompt the child to take a turn or do another action. They can also help to un-stick the activity, but without the disruption of re-starting the entire game (e.g. reset Andy to his normal image and actions). The butterfly hot key was added as a result of testing with typically developing children, and was meant to provide an on-demand novel aspect that might attract children's attention back to the screen, or re-engage them if the researcher encourages them to attend to the screen.

7.5 Altered versions of mini-games, with DDOs

7.5.1 Overview of altered game development

The chapter introduction explained that the *Andy's Garden* suite of games consists of a basic game version, in which everything has its normal behaviour as a baseline for child expectations, and two altered versions that each introduce a different set of DDOs. The intention was for all evaluation study participants to play all game versions. This section discusses the *games session*, or the content a child plays on a particular day, as a key unit of planning. It outlines the choice and distribution of DDOs between the altered game versions.

DDOs need to be considered at both activity and environment levels. A group of DDOs might work well in a particular activity, but create an imbalanced experience when encountered together with the DDOs in the session's other activities. Alternately, the session might fail to provide adequate DDO variety to engage all children. Creation of the DDOs for the altered mini-games and sessions went through five steps:

1. **Generation** of DDO ideas, extrapolating from ECHOES discrepant aspects, and

Key	Action
Shift + click	Use a touch-disabled menu button; Move touch-disabled apple if Andy turn interrupted
N	Normal: Manually return Andy to his default image and cancel any actions in progress
M	Menu: Manually return to the menu at any point; End carrot game, go to "goodbye" sequence
R	Repeat instruction: Andy repeats activity instruction, says "your turn"
A	Andy take turn: Andy says "my turn" and takes turn immediately
S	Stars reward: Manually play stars reward after short pause, ends activity
T	Thank you: Andy immediately says "thank you" and smiles
W	Wow: Andy immediately says "wow" and smiles
C	Carrot prompt: Andy asks "Can you pick a carrot?" [CARROTS ONLY]
U	Andy non-action: Andy says "my turn", does not take any actions [ALTERED ONLY]
B	Butterfly: Triggers butterfly to fly across screen [ALTERED ONLY]

Table 7.1: List of researcher hot-keys for additional system control.

possible changes to elements of the basic game versions.

2. **Screening** of DDO ideas for technical feasibility, and potential problems with respect to resolvability.

3. **Feature labelling** for each candidate DDO (e.g. aspects involved, interactivity).

4. **Provisional allocation** of DDOs to the six altered activity versions, based on the DDO features.

5. **Balancing** the DDOs allocated to each activity, in light of other DDOs in the same session. This was a period of adjustment, trading DDOs between activities and sessions until it seemed like each activity and session offered a balanced variety of DDOs, at a similar frequency.

It would be possible to generate a near-infinite number of plausible DDOs for *Andy's Garden*, even when limited to garden-relevant objects and when respecting the other constraints. It is important to emphasise at the outset that the DDOs that were eventually chosen are *one* possible set, meant as a tool to investigate discrepancy. The final list of DDOs and their allocation is reported at the end of the altered games section, with descriptions and images.

7.5.2 Designing to accommodate planned evaluation study

Andy's Garden uses the *games session* as the main unit of content and planning, which is to say the amount of content that a child could feasibly play through during a single evaluation session. A games session was estimated to be about 15 minutes worth of continuous play (based on previous studies with this age and ability group). The evaluation study aimed to collect about 45 minutes of game play video for each participant, which would equate to three sessions (on three separate days). Sessions are separate experiences, that *each* need to engage the child and provide sameness and safety. The task of choosing and organising possible DDOs must thus also be considered at the session level, as well as the individual activity level.

Unlike ECHOES, which was fairly flexible about the content to be presented on any given day (due to the far longer evaluation period), it is now very important that children see content in a certain order. The integrity principle (Chapter 6) highlights that in order for a designer to violate child expectations, s/he must give an opportunity for children to *establish* expectations. They must encounter "baseline" content or

Session 1	**Session 2**	**Session 3**
Basic Flowers	Altered Flowers 1	Altered Flowers 2
Basic Apples	Altered Apples 1	Altered Apples 2
Basic Carrots	Altered Carrots 1	Altered Carrots 2

Table 7.2: Game versions to be played in each *Andy's Garden* evaluation session

behaviours for some period of time, in order for them to have the body of knowledge and expectations that would enable them to recognise DDOs as discrepant. In *Andy's Garden,* the basic games versions provide this baseline experience. After a child is familiar with these, s/he would be in a position to detect discrepancies.

Mapping the planned volume of content to games sessions would equate to one session for baseline games versions, and one session for each set of altered versions, which include DDOs. This distribution represented in Table 7.2, with each child seeing all three mini-games in each session, rather than focusing each session on a separate activity. There are two main reasons for this choice. The main one is to follow the already-successful model of ECHOES sessions, in which children played a mixture of learning activities on any given day. Once children grasped an activity's goals, actions and patterns, they appeared to have no trouble remembering these between sessions. Secondly, playing different mini-games appears more likely to keep children engaged over a whole session and feeling positive about their experience. In ECHOES, moving on to a "different game with Andy" often helped re-engage and excite children if they were discouraged about a particular game, or simply preferred another one. It was predicted that children might react negatively to being required to play the *same* mini-game multiple times back-to-back, without changes or choices. Especially if the child had difficulty with a mini-game (or preferred the content in the *previous* session), it is a real risk that he might refuse to continue playing all the versions scheduled for a session—meaning that the missing data would be from the altered game versions, the main focus of the study. Presenting all three mini-games in each session seems unlikely to cause any problems, and may result in a more positive child experience.

7.5.3 Expectation formation and DDOs

Before discussing DDO choice and placement, it is worth returning to the idea of novelty and expectation-formation. The annotation scheme detailed in Chapter 4 used a heuristic "rule of three" to judge the point at which a child should be assumed to have developed expectations (see Section 4.5.3.3). In terms of design rather than annotation, the key points would be:

- A child is assumed to have formed an expectation about an aspect if s/he has been exposed to it three times.

- Exposure is more specific than an aspect being present. It means that the child might reasonably have been able to attend to it and notice it (e.g. the researcher or child were not talking over the top of the new sound effect, the child was looking at the screen when Andy took his turn, etc.). In some cases, exposure will mean a child's direct action, such as learning about the magic cloud by trying it out.

In the final annotation of Study 1 data, this heuristic seemed to be a good fit for most children's behaviour. This has now been borrowed as a good heuristic for design: that children should ideally be exposed to items, behaviours, or patterns *at least three times* before these are affected by a DDO, because otherwise they may not yet have clear expectations in place. The rule of three also suggests that we could use a potentially-surprising DDO twice or three times, before it effectively creates a new expectation. This heuristic has not been followed in all cases with respect to exposure, due to the short number of sessions, but has been respected wherever possible. It is here that having two near-identical growing activities becomes extremely useful, because the child gets two exposures to patterns (and actions, behaviours, etc.) in each session, even without re-playing any of the games. It is true that this creates a disparity in which there are fewer exposures to behaviours specific to the apple game. This is also useful, because it is a further opportunity to investigate expectation-formation: how many exposures are sufficient? It might be fewer than currently thought. Of course, many behaviours and patterns are common to all three games (e.g. Andy's timing and type of feedback).

7.5.4 DDO Allocation process

7.5.4.1 Generation and screening of DDOs

A number of candidate DDOs were brainstormed while creating the initial game concepts, and more of them borrowed directly or extrapolated from the ECHOES aspect list. For example, Andy making a ball-sorting mistake suggests Andy making an analogous apple-sorting mistake. In some cases, children's *perceived* discrepancies in ECHOES regarding unusually-sized, new, or missing items were proposed as *actual* changes here. Others candidate DDOs were generated simply by considering how objects or effects in the basic game version could be changed. For example, the cloud could change its appearance or size; it could produce different rain, more rain, or something other than rain; it might make a new sound, or no sound. Perhaps instead the cloud disappears, has a sunshine attached, or appears as two tiny clouds. Listing *some* of the possibilities for the cloud starts to illustrate the vast number of potential options, even before any objectively new objects or effects are introduced.

Generating candidate DDOs was not meant to be systematic or exhaustive, only to generate a large enough pool of ideas from which to start narrowing down to a final set. The first step is screening candidate DDOs for potential problems:

- Apparent threats to integrity, at the activity or environment level (e.g. apparently alters an activity goal, may create confusion between activities)
- Affecting (or appearing to affect) the child meeting an activity goal: could these be made resolvable? How?
- Potential to confuse or upset a child, potential for otherwise creating a negative game experience

Some candidate DDOs were deemed borderline problematic, but left in the set because they were potentially useful research-wise. For example, Andy suddenly vanishing at the end of an activity was thought to be potentially too abrupt. It might potentially confuse children when he then appeared again at the start of the new activity.

At this stage, some DDOs were also ruled out due to technical complexity. For example, replacing one cloud with two would have meant substantial behind-the-scenes reconfiguration of variables and object behaviours. As there are plenty of other cloud-related DDO options, this was struck off the list. Other DDOs fell afoul of the requirement for resolvability (i.e. no obvious or technically feasible way to do this), or were too similar to other options (did not offer increased variety).

7.5.4.2 DDO feature labelling

To move toward allocating DDOs to activities, the options were each labelled according to their features. Those features considered relevant to *Andy's Garden* included:

1. Activity: the mini-game(s) in which the DDO could potentially be included.
2. Activity phase: Where or when in an activity the DDO would be apparent (i.e. immediately when activity loads, during main play, during ending sequence).
3. Which aspects are affected (e.g. which object, event timing, action...).
4. How “big” a DDO this is expected to be. Does it seem quite subtle, or attention-grabbing?
5. Whether Andy is involved or affected in the DDO.
6. Interactivity of the DDO (active or passive).
7. Predicted subtype for that DDO (novelty, surprise, non-event).
8. Modality of the DDO (e.g. sound, visual, both).
9. Relationship to ECHOES aspect groups (if relevant).

Some of these features are useful for the designer, such as discrepancy subtype, but less meaningful in terms of the child’s experience and understanding of the games. Planning in *Andy’s Garden* prioritised those features that described the child’s direct experience.

It is an important distinction that label seven is *predicted* discrepancy subtype. As discussed in earlier chapters, discrepancy is about subjectively perceived inconsistency. For planning purposes, the designer can make a prediction about how a child may perceive a DDO, based on what s/he thinks the most relevant part of the DDO may be. This can be surprisingly tricky. An example may be useful here. Consider a child action which always results in sound effect X. Now, a DDO causes that action to result in previously unheard sound effect Y. The designer might predict that this would be a surprising event—“hey, this is different!” However, for the child it might be more relevant that effect Y is novel, or that effect X is *missing*, a non-event. This is a particularly difficult example because it includes both an objectively novel element and a change. Most DDOs include only one or the other. In many cases, child reactions will give a cue to how s/he perceived the DDO. In cases where the reaction does *not* give this cue, the annotator must decide how to label a discrepancy, in light of a child’s prior system interaction (i.e. all the aspects involved may definitely be familiar, ruling out novelty). Here, how *other* children have reacted to the DDO may be a good guide to labelling, especially if this has contradicted the designer’s predicted label.

7.5.4.3 Provisional allocation to activities

Allocation followed several heuristics, applied in a particular order. The unpredictable nature and random distribution of the "naturally occurring" discrepancies in ECHOES does not provide any particularly good guidance here. These heuristics represent a mixture of common sense and extrapolation from the integrity principle.

1. Within an activity, the same one or two aspects should not be affected repeatedly but differently, while others are unaffected. It appears better to "spread out" the DDOs across activity aspects.

2. DDOs should also be spread out in time (i.e. over activity phase). They especially should not cluster in the beginning or ending of a mini-game, because these are an important source of sameness across activities. The majority of DDOs should probably be in the longer (and already more variable) middle part of the mini-game.

3. To keep interactions manageable and preserve the integrity of cause-effect relationships as much as possible, DDOs should not compound one another or cancel each other out. In other words, no more than one DDO should affect the same object at the same time.

4. There probably should not be more than one "big", attention-grabbing DDO in an activity. This might become disruptive. There also should probably not be more than one completely novel object or behaviour (given the short length of the activities), or more than one "persistent" DDO that remains visible/ in effect for the entire activity.

Here, activity phase was used as the starting heuristic: that no more than one DDO should be *immediately apparent* when an activity opens. In the final version, a few activities did end up with multiple opening DDOs, some of which were subtle changes to the background. After starting and ending phase DDOs were provisionally chosen for the six altered activities, a number of "middle" DDOs were chosen from the pool, and shuffled around to give each activity a spread of DDOs for different aspects and modalities.

A key constraint on the initial allocation process was expectation-formation. At the end of a session with the three basic mini-games—assuming these were played only once—what exactly would the child have seen, and what might s/he expect? This was particularly important for the apple game, as much material was already repeated in the growing game. Due to the short overall study, some aspects had to be altered after only two potential exposures, even though this was less than ideal[10].

7.5.4.4 Balancing DDOs across sessions and activities

The balancing process sought to roughly equalise the number of DDOs in each session, ensure variety, and make sure that certain types of DDOs would be represented in the altered game versions. Predicted discrepancy subtypes were used as a balancing criterion after there was a semi-final set for each activity. This consideration led to substituting DDOs across activities (and from the few unallocated ideas) to achieve a rough balance of the three subtypes. For example, a predicted non-event that affected the cloud might be switched for a surprising event affecting the cloud. The semi-final list was similarly checked and adjusted for activity and passivity of DDOs, to ensure both were adequately represented. Balancing also tried to give each altered activity at least 1-2 DDOs involving Andy, and at least one involving sound.

Novel and non-event DDOs are, by numbers, deliberately under-represented in the designed DDOs. The main source of novelty within *Andy's Garden* is the *inherent novelty* of encountering digital objects and discovering game play actions for the first time. Session 1 (the basic game versions) is predominantly about experiencing novelty. In the altered games, several completely novel aspects (snail, butterfly, rainbow, sounds; see 7.3) are introduced to test the prediction that these should be perceptible as new or different, when surrounded by now-familiar elements. Non-events are also under-represented because, in predicting the type of interface and system errors that are likely to occur, most of these might be experienced as non events. For example, a known error from testing is that Andy may start to trigger his turn taking, saying "my turn", but then he stops and nothing happens. Troubles with the touch-screen apparently being broken or "not working" are similar opportunities for non-event-type discrepancy. While there *are* designed non-events, there are fewer than these of designed events. Non-events are likely to occur anyway due to child perception and/or genuine error during play, so fewer of them have been added to the game design.

[10] Though as it turned out in the evaluation study, most participants played the basic games versions twice during session 1 and seemed to form expectations quickly.

7.5.5 Final DDOs in Andy's Garden

7.5.5.1 Description of DDOs

The altered versions were created by copying the basic game versions, then editing these to introduce the DDOs. The final list of DDOs, with some of their key features (with respect to design decisions) are given in Table 7.3. The DDOs have been named to reflect their nature and affected aspects as much as possible. Those DDOs than involve new or changed visual elements are illustrated together in Figure 7.18. A key to the images is given in the caption. The following section presents the DDOs organised by activity and session, with some commentary on how many times each would be seen. All DDOs have a brief description in Appendix G. A few are selected for further discussion here, as they illustrate some of the decision-making process of developing and implementing DDOs.

A self-resolving missing object Children playing ECHOES sometimes thought objects were missing (when in fact they were part of another game). Here, objects are actually subtracted to see how children react. When a growing activity loads, the cloud is absent instead of being in the upper left of the screen. The cloud is necessary to meeting the activity goal; leaving it off-screen until the child does something would be creating a "bottleneck" or demand for specific action, of the type we wish to avoid per the design principles in Chapter 6. After some seconds (a noticeable pause), this DDO *self-resolves* when the cloud glides onto the screen to its usual start position. Whether the child does anything or not (and whether he notices or not), he will still be able to complete the activity.

Andy's mistakes Andy's mistakes were amusing and popular for most children in Study 1. Because he does things correctly most of the time, the mistakes violate both his usual pattern of activity and role as the child's knowledgeable guide. Here, Andy can make two kinds of mistakes in the apple game: he can throw an apple to the wrong colour tube, or throw and miss the tubes entirely. Either way the apple ends up bouncing on the ground. This DDO also self-resolves, in that it returns the environment to its original state of apples on the ground, waiting to be sorted. For both types of throwing mistakes, Andy is set to make a mistake on his first turn of an activity. Thereafter, a random number generates the trajectory he will use. In sessions 2 and 3, Andy has a 50% chance of making a mistake on any throw after the first (i.e. miss tube

Phase	DDO	Interactivity	Activity	Predicted subtype	Session
Start	Changed growing behaviour: Partly grown at start	Passive	Growing	Surprise	3
	Cloud missing at activity start	Both	Growing	Surprise	2,3
	Fence removed from garden	Passive	Growing	Non-event	3
	Planter colour changed	Passive	Growing	Surprise	2
	Snail added to garden	Both	Growing	Surprise/ Novelty	3
	Some apples do not fall from tree	Both	Apples	Surprise	3
	Some apples giant-sized	Passive	Apples	Surprise	2,3
Main play	Andy apple mistake: throws to wrong tube	Passive	Apples	Surprise	3
	Andy apple mistake: throws and misses tube	Passive	Apples	Surprise	2,3
	Andy mis-statement	Passive	Multiple	Novelty	2
	Andy new feedback: You did it!	Passive	Multiple	Novelty	2,3
	Butterfly flies across screen	Passive	Multiple	Surprise/ Novelty	3
	Carrot turns into strawberry when picked	Active	Carrots	Surprise	2
	Changed growing behaviour: Plant grows in 1 rain	Passive	Growing	Non-event	(2,3)
	Cloud sound changed: Thunder	Passive	Growing	Surprise	3
	Cloud sound missing	Passive	Growing	Non-event	2
	Fanfare changed to novel sound	Passive	Growing	Novelty/ Surprise	2,3
	Flowers: All different	Active	Flowers	Non-event	3
	Flowers: One different	Active	Flowers	Surprise	2
	Giant raindrops	Active	Growing	Surprise	3
	Some carrots giant-sized when picked	Active	Carrots	Surprise	2
	Tube sound changed to novel sound	Passive	Apples	Surprise	2,3
	Tube sound missing	Passive	Apples	Surprise	3
	White tube makes 2 apples	Passive	Apples	Novelty	3
	White tube makes strawberry	Passive	Apples	Novelty	2
End	Andy changes end-of-activity utterance	Passive	Multiple	Surprise/ Non-event	3
	Andy disappears at end of activity	Passive	Flowers	Surprise	2
	Fireworks reward sound changed: magic noise	Passive	Multiple	Surprise	2
	Rainbow appears at activity end	Passive	Carrots	Novelty/ Surprise	3

Table 7.3: List of designed DDOs with their key features and the session(s) in which they appear.

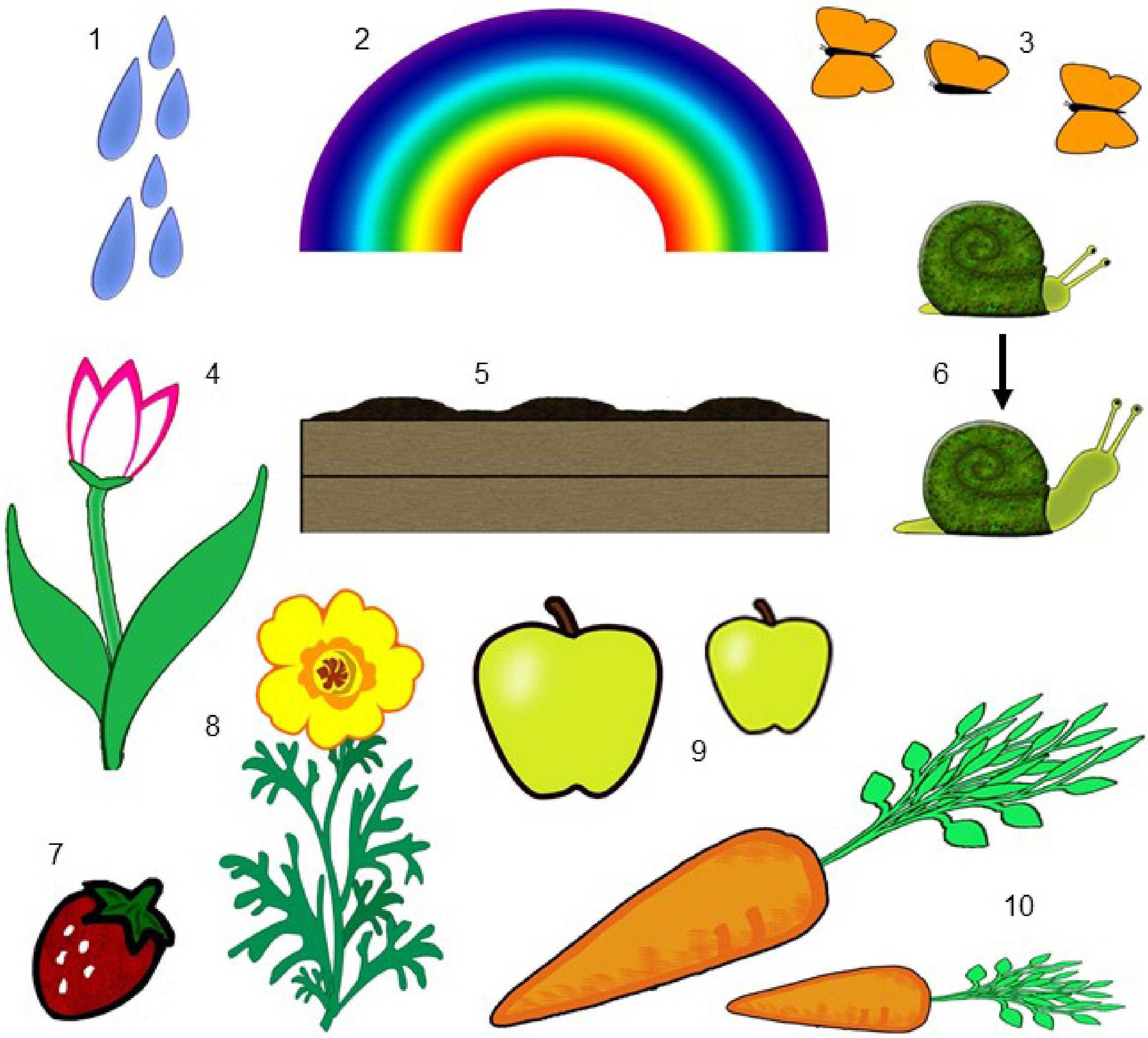


Figure 7.18: Items changed or introduced to altered games versions. 1) Giant raindrops 2) Rainbow 3) Butterfly, flapping 4) Pink-and-white tulip 5) Light-coloured planter box 6) Snail in 'in' and 'out' positions 7) Strawberry 8) Yellow poppy 9) Giant apple, with regular apple for scale 10) Giant carrot, with regular carrot for scale.

or throw to opposite colour tube). The design intention is for each child to see these mistakes at least once, though (as in ECHOES) children are more likely to see zero mistakes than many mistakes, if they cannot inhibit their own action for Andy to take a turn.

Language-based DDOs The original ECHOES aspect list had several intriguing instances of DDOs that could be considered language-based, or about an inconsistency between a partner's utterance and the environment[11]. Only a few children reacted to events of this type. Given the social importance of language and the challenge it frequently presents to children with ASC, this seemed like an important lead on which to follow up from ECHOES.

Andy's Garden includes several DDOs in which Andy's utterances conflict with the state of the environment in some way. At least once in each activity, he says something incorrect (e.g. that a yellow flower is a red flower, that a strawberry is a carrot) or irrelevant (e.g. "I like apples!" when he is rained on in the growing game). He does not introduce any new phrases that he then uses incorrectly, he only substitutes phrases he has already used elsewhere. This DDO is set to happen in response to specific events (e.g. completing a particular flower) rather than choosing between phrases and randomly being wrong some of the time. There are several additional DDOs that change or add to Andy's language, but in a way that is congruent with the environment. For example, the addition of a new feedback option ("You did it!") and changes to his end-of-activity utterances both add more narrative about what the flowers and apples are for.

Violating expectations from outside the environment Per the ECHOES aspect list, children in Study 1 reacted to few things that appeared to represent a violation of rules from outside the environment. All of them related to the basket, which did not have the solid, container-like properties of a real basket. Other parts of ECHOES that violated real world rules *were explicitly identified as "magic"*, such as swiping a flower to turn it into a bouncy ball. *Andy's Garden* has several DDOs that obviously violate real world rules, some of which are explicitly described as magic and some of which are not. For example, the "magic" white tube (so-called in interaction between the researcher and child player) is altered to produce *two* apples of an opposite colour to the input apple

[11]For example, aspect label 16.2 *Andy said the ball changed colour but pointed to one that had not*, or 16.5 *Researcher said Andy would take his turn or do a certain action, but he did not* (see other examples in AG 16, in Appendix E)

(i.e. violating conservation of number). It also turns an apple into a strawberry. In the carrot game, there is a similar DDO which is *not* announced as magic, in which a carrot suddenly becomes a strawberry when it is picked.

7.5.5.2 DDO allocation

The final allocation of DDOs, in six altered games across two sessions, resulted in an estimated 26-31 designed DDOs per session. This can only be an approximate number because multiple DDOs depend on the child's action. Andy cannot miss his apple turn if the child will not let Andy take a turn! Or, a DDO that relies on a timer or counts of actions may not be experienced if a child completes a game very quickly, with a small number of actions. Tables 7.4 and 7.5 list the number of times that various DDOs appear in each session and mini-game. Asterisks indicate those DDOs that may vary in the number of times they are encountered; the number listed is an estimated minimum. The other DDOs should appear a set number of times.

The estimate of DDOs per activity and session does not include any inherently novel aspects, and does not include an estimate of further DDOs related to individual children's perception or to touch screen issues. It is really not possible to estimate these beforehand, even though we can be almost certain that children will have some DR pairs with these causes. While we can enumerate different objects that will be objectively novel to the child when playing the basic game versions, novelty can also include patterns, and relations between objects, or results of child actions. The answer to "how many?" can change, depending on the level of detail at which one looks, which actions or effects are considered as "units" or sequences, versus in terms of their separate steps or component modalities.

As a final note, some DDOs are introduced at a certain point and then remain in all later activities (e.g. Andy's new feedback phrase "You did it!"). For the purposes of design, it was assumed that these would become familiar after a certain point, if children noticed them at all. Only their first three exposures (actual or predicted) are listed in Tables 7.4 and 7.5, per the "rule of three" heuristic (see Section 7.5.3). This is not the same as a DDO like *Andy mis-statement*, which is listed as a DDO in all activities because he uses *different* mis-statements in every activity.

Phase	DDO	Flowers	Apples	Carrots
Start	Cloud missing at activity start	1		
	Snail added to garden			1
	Some apples giant-sized		1	
Main play	Andy apple mistake: throws and misses tube		1*	
	Andy mis-statement	1	1	1
	Andy new feedback: You did it!	2*	1*	
	Butterfly flies across screen		2	
	Changed growing behaviour: Plant grows in 1 rain			1
	Cloud sound missing			2*
	Fanfare changed to novel sound (bike horn)	1		
	Flowers: One different	1		
	Giant raindrops	2*		
	Some carrots giant-sized when picked			2
	Tube sound changed to novel sound		2	
	White tube makes 2 apples		1*	
End	Andy changes end-of-activity utterance	1		
	Fireworks reward sound changed: magic noise			1
Total DDO estimate per activity		9	9	8
Total DDO estimate for session 2			26	

Table 7.4: DDOs implemented in altered game version 1 (mini-games played in session 2). Counts marked with an asterisk (*) indicate DDOs that are intermittent (based on action counts, timers, or randomisation) or depend on child action in such a way that they may appear a variable number of times.

Phase	DDO	Flowers	Apples	Carrots
Start	Cloud missing at activity start			1
	Fence removed from garden	1		
	Planter colour changed			1
	Snail added to garden	1		1
	Some apples do not fall from tree		1	
Main play	Andy apple mistake: throws and misses tube		1	
	Andy apple mistake: throws to wrong tube		1*	
	Andy mis-statement	1	1	1
	Butterfly flies across screen	2		
	Carrot turns into strawberry when picked			2
	Changed growing behaviour: Plant grows in 1 rain	1		1
	Cloud sound changed: Thunder	3*		
	Fanfare changed to novel sound	1		
	Flowers: All different	1		
	Giant raindrops			2*
	Tube sound missing		3	
	White tube makes strawberry		1*	
End	Andy changes end-of-activity utterance		1	
	Andy disappears at end of activity	1		
	Rainbow appears at activity end			1
Total DDO estimate per activity		12	9	10
Total DDO estimate for session 3			31	

Table 7.5: DDOs implemented in altered game version 2 (mini-games played in session 3). Counts marked with an asterisk (*) indicate DDOs that are intermittent (based on action counts, timers, or randomisation) or depend in child action in such a way that they may appear a variable number of times.

7.6 Usability testing with typically developing children

7.6.1 Participants and procedure

Semifinal versions of the basic and altered games versions were tested with six young *typically developing* (TD) children[12], aged 4-6 years (mean 4.8; M=5, F=1). The main goal of these sessions was usability testing of *Andy's Garden*, and to get a sense of how children were playing the games: how long did they take to do each mini-game? What things did they find particularly interesting, or not? Observations of children's game play and informal discussion were used to answer these questions.

Children were recruited through an internal mailing list at the University of Edinburgh School of Informatics, and were all relatives of staff members. Parents and children were provided with information and consent forms similar to those used in the evaluation study (Chapter 8, see forms in Appendix H), except that no additional measures were being collected. It was made clear that child participation was for the purposes of improving research software, was not a test of child abilities, and would not be formally analysed.

Testing sessions lasted 30-45 minutes each and were conducted in a small, quiet room at the University of Edinburgh. Parents (and sometimes siblings) remained in the room during most of the sessions, out of the way unless the child spontaneously initiated to them. Five out of six children agreed for their game play and critique to be video-recorded for later review; the sixth child's session was documented via note-taking. As it was initially unclear how long game play might take, each child was scheduled to play through the basic versions of two mini-games, and one altered version of those mini-games. However, in testing, all children played the basic version of all three mini-games and at least one altered version for each. They played more quickly than anticipated, with older children playing the fastest. It took the six-year-olds about 10 minutes for three mini-games, but took the four-year-olds about 15 minutes for the same content. Mixed in during and after their game play, children chatted to the researcher (and sometimes to their parent) about the games. Where possible, the researcher specifically probed for likes, dislikes, and suggestions about other things the garden might do. Some of these ideas resulted in design changes, discussed below.

[12] Children without ASC or other developmental disabilities.

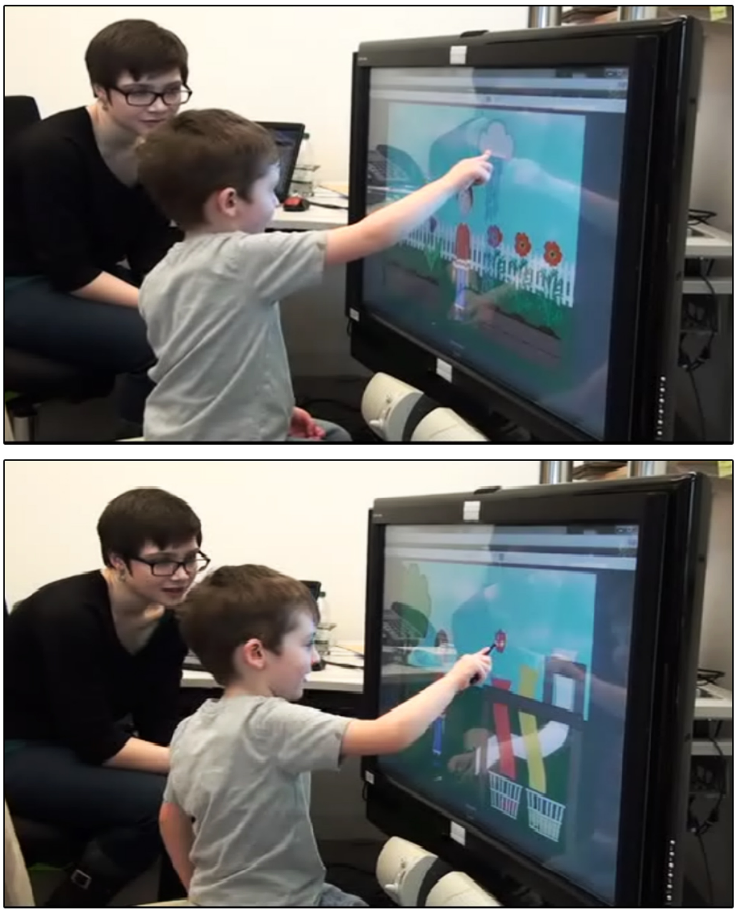


Figure 7.19: A typically developing child testing semi-final versions of the flower game *(top)* and apple game *(bottom)* on the large (42”) multi-touch screen for which the games were originally designed.

7.6.2 Observations from TD testing

The TD testers clearly enjoyed playing the games, and were willing to experiment and try out as many actions as possible. Even the six-year-olds, who found the games fairly easy, remained fully engaged for a whole session. Unfortunately, changes to object physics did not eradicate some of the same touch screen troubles seen in ECHOES. Children still struggled to maintain consistent pressure, and often dropped objects. Regardless of these problems, they showed high motivation to continue play and complete activities, even when they had to try many times.

In observing children play, it appeared that many of the designed DDOs successfully attracted children's attention and comment. Two children were particularly enamoured of the apple changing into a strawberry, which helped make apples "the best game". Many of the sound-related and growing-behaviour-related DDOs were apparently missed by all children, and were perhaps too subtle. A decision was taken *not* to change these and make them more noticeable, due to the wide variety in child interests (and attention to detail) seen in ECHOES. More importantly, a goal of the games as a tool is to help investigate the range of DDOs that work to attract child attention and motivate communication. Having unsuccessful, too-subtle examples is also useful to understanding discrepancy.

As in ECHOES, the children often struggled to self-inhibit during Andy's turn-taking, regardless of prompting. Some children did not see him demonstrate actions because they *could not* wait for them. Their interference with his turn-taking tended to cause problems, such as Andy freezing because he could not execute two actions at once. This required some changes to object properties, listed in the following section.

The children who mentioned it did not much like Andy's expanding arms in the growing games. Like the expert design critique participants, they found it rather off-putting. One tester, age 4, explained that he didn't like the "zig-zag" arm because "it could have fallen off and made everything red" (a reference to bleeding?). This was somewhat worrying, but unfortunately, it was not possible to change Andy's turn-taking because there was not an alternative idea ready to implement. Moreover, a new turn-taking implementation seemed unlikely to be ready in the short time available before the evaluation study.

7.6.3 Changes resulting from TD testing

The testing sessions were spread out over a week, and there were many minor changes made between children—especially to address small bugs, tweak timings, and adjust the sizes and physics of objects for ease of interaction. Children's play and informal critique identified some larger and more desirable changes, many of them around interaction with Andy. Unfortunately, some child suggestions were outside the scope of the current design because they would have altered the mini-game goals or made them more difficult to play. Other suggestions (e.g. for Andy to grow/shrink or be wet in the rain) echo those of the expert design critique. These have obvious appeal but cannot be easily implemented in GameSalad, or would conflict with the implementation Andy's other behaviours.

The child suggestions and observations that were feasible and could be incorporated into the design were introduced at several different points in the development, as they became ready. For simplicity, the changes are all reported together, below, and not in an exact chronological order.

Larger apples and slower physics On the large multi-touch screen, all children still had trouble with manipulating the apples and carrots. Gravity and dropped objects were still a problem. Consequently, the object physics were slowed, and apples enlarged. These factors were adjusted again several times over the testing period.

Make Andy respond when his body is touched The TD children wanted Andy to do more, echoing comments from the expert design critique. Their suggestions for Andy to jump or dance could not be implemented, but Andy was changed to say "hello" when his face was touched. As with "it's raining on me!", Andy will only say this up to a maximum number of times and then stop responding.

Make Andy eat carrots One child specifically suggested that Andy should eat the carrots, and another really wanted the carrots to do something or go somewhere once they had been picked. The implementation of this change is discussed further below.

Exaggerate end-of-activity reward Children were unimpressed by the original end-of-activity reward, and did not find it very rewarding. It was originally more pastel than the one shown in this chapter. It also originally had a more twinkly, "magical" noise which was easily drowned out, especially if the child talked at the end of the activity. One suggested that "rainbow fireworks" would be much better. Thus, the animation was made larger, brighter, and given a fireworks-like sound effect.

Child-proof Andy's apple-sorting As noted above, children struggled to wait and watch Andy's apple-sorting. When they tried to pull apples away from his hand or off their thrown trajectory, this created errors. The apple that Andy throws (Figure 7.16) is now a different type of object (in GameSalad), with different properties than the normal apples. The "throwing apples" were changed so as to be unresponsive to touch *until* they entered a certain area of the screen. The child cannot affect the turn, but in the DDOs where Andy throws and misses, the apple becomes touch-responsive *after* it has missed the tubes. The child can then pick it up and use it normally.

Add butterfly hot key Children's attention did stray from the screen, even though overall engagement was high. It seemed useful to have an "on-demand" event as an additional tool for the researcher, to help re-capture children's attention on the screen or break a long pause where nothing was happening.

One child's specific idea that he wanted Andy to eat the carrots turned out to be an invaluable design contribution with respect to making Andy more interactive and engaging. Andy was modified to react when carrots were held over his face. To avoid accidental eating if children were moving carrots around the screen, items must be held over his face for several seconds before he says "Mmm!" and makes a crunching, chewing noise as he eats the first bite. After another few seconds, he says "Mmm!", crunches some more, and the item disappears (as in Figure 7.20). To ensure that Andy's properties were consistent across mini-games (i.e. that he would behave with *integrity*), he is also able to eat apples, and the strawberry when it appears. Just for fun, the snail in the altered game versions was also updated so that he can eat carrots and strawberries, using the same mechanics with a different sound effect. Andy eating carrots became one of the most popular basic game actions during the evaluation study.

7.6.4 Emergency pre-evaluation study changes

After TD testing but prior to evaluating *Andy's Garden*, there was a last-minute change of equipment when the 42" touch screen was damaged in transit. The new screen was

Figure 7.20: Per a request from child testers, Andy can now eat the carrots, apples, and strawberries.

a 22” touch screen monitor, without multi-touch capability. The switch from multi-touch to single touch did not affect the design, but a number of objects were enlarged and their physics slowed even further, to make them more manageable on the smaller screen. The mini-game images presented earlier in this chapter show final object dimensions (except for those images that are specifically labelled as referring to early game versions).

7.7 Reflection on high-level principles in the current design

Having presented basic and altered versions of each mini-game and some outcomes of child usability testing, we now step back to briefly consider the *Andy’s Garden* design as a whole, and how its design choices have sought to implement the concept of *motivating but manageable design* with DDOs (introduced in Chapter 6). This is done in two parts: considering sameness in and around the games, and considering how the high-level design principles are reflected in the design.

7.7.1 Sources of sameness in Andy’s Garden

Reflection on Study 1 led to the working hypothesis that substantial sameness can counterbalance discrepancy, making children’s overall experience manageable. The *integrity* principle is effectively another working hypothesis, that says that part of this sameness is in the form of establishing lawful behaviour, patterns, and expectations at environment and activity levels. The task of designing DDOs in *Andy’s Garden* has thus also been a task of designing sameness, *as well as* designing differences that may be perceived as discrepant. This is an important point. Seeking to design motivating situations (i.e. DDOs) *that are individually manageable*, is not the same as actively designing for both motivation *and* sameness across an entire environment. For example, DDOs may be individually manageable, but overwhelming in concert.

Reading a description of the mini-games gives a poor sense of their patterns and degree of repetition, compared to what children actually experience in playing the games. Some of the key sources of sameness are thus identified here:

- Around the technology, routines in technology use
  - A games session has a clear internal pattern of playing, and returning to the menu, and playing again. It always gives clear warning of what is next. Children always played flowers, apples, and then carrots, with time at the end to re-play the game if children requested it.
  - In the evaluation study (Chapter 8) sessions were always in the same place, working with the same researcher.
- Material around mini-games
  - The menus (i.e. sequence of 3 menus) never vary.
  - Andy's open/close sequences are repeated in every session, and never vary.
- Material shared or held constant across mini-games
  - Despite changes to objects, the garden settings are always recognisably the same.
  - Very similar opening routines, in which Andy uses one phrase to announce the goal, then demonstrates an action. At the end of each mini-game, there is the combination of the same visual reward, and Andy's announcement of task completion.
  - Both the growing and sorting games have intermediate rewards and positive feedback (completed plants, sorted apples), using the same fanfare sound (though this was occasionally altered for DDOs).
  - Andy shares many of his actions and utterances across the game (e.g. feedback phrases, "my turn", "finished").
  - Children's direct actions within the game have a high level of repetition: there are not many *different* things for them to learn. Both sorting the

apples and growing plants require one main type of activity (move apple, hold cloud touch) and then this is repeated multiple times.

- The mini-games never vary in their goals, level of difficulty, or scale (i.e. number of objects to grow/sort). Children never have to do more than they might expect, or (potentially) feel cheated because there is less.

Taken together, *Andy's Garden* has many sources of sameness, and may even surpass the amount of sameness in ECHOES, given the smaller set of activities and Andy's more limited behaviour. The unchanged elements, routines, and repetitions are thought to provide an "insulating" effect, such that children will feel safe when they encounter DDOs. The working hypothesis of motivating manageability suggests that that sameness and safety are a firm footing, from which children can explore and positively interact with (and about) discrepant aspects.

7.7.2 High-level principles in the current design

Each of the six principles delineated in Chapter 6 describes a factor thought to be important in terms of facilitating positive actions and communication around discrepancy. Each principle serves as a working hypothesis, under the overall concept of a motivating but manageable environment that balances out DDOs with sameness. At their point of presentation, it was emphasised that motivating manageability was thought to be an *overall effect* of the factors represented by each principle. All of the principles are reflected in the current design, and are interlinked to some extent. Brief comments on specific principles are presented below.

INTEGRITY The first games session (basic versions) is a dedicated opportunity for children to explore and develop expectations about *Andy's Garden*. This period of expectation-establishment and the degree of sameness in the environment is meant to enable children to recognise later-session DDOs as being new or different in some way. As noted above, there are many sources of sameness and repetition within each mini-game and session, that help to establish the garden as an ordered, predictable place. This also should make clear to children that all three mini-games are part of the *same* environment, and thus the same rules apply. At an activity level, the visual-schedule-like menu, unchanging activity order and Andy's opening utterances all make clear which activity is being played at any time. Where objects re-appear across activities, they have consistent properties *unless* these are altered in a DDO.

FLEXIBILITY Both the growing games and the apple sorting game offer children some flexibility of action. All the plants must be grown and all the apples tidied away, but the child can do these in any order, with as much stopping and starting as s/he wishes. With growing, the child can also switch back and forth between plants, instead of having to finish one before going on to the next. The apple and carrot games give options for non-goal activity, which is another source of flexibility. The apples have capacity for play (bounce, throw, stack) outside of sorting, and the picked carrots also do to some extent. The white apple tube provides interesting effects, but is not directly a part of the sorting goal. The child can also—potentially—mix picking carrots and playing with them, or feeding them to Andy, with further use of the cloud.

As discussed in relation to the original ECHOES, not forcing children to interact with (potentially) discrepant aspects is thought to be important to an emotionally manageable experience. As much as possible, DDOs affect only one object at a time (excluding changes to the background), leaving the rest of an activity "intact" and operating as usual. Newly introduced aspects, such as the snail, are not part of the activity goal and can be ignored or investigated as the child prefers. To some extent, flexibility is about giving children choice and ownership during game play.

RESOLVABILITY Many DDOs in *Andy's Garden* do not require resolution, because they do not interfere with or prevent a child's actions. For example, changing sound effects and flower colours are tangential to the activity goals, as is introducing a snail or a rainbow that are independent of everything else in the activity. Some others self-resolve, returning the environment to its original (or an equivalent) state afterwards regardless of whether the child notices or takes any actions, such as when Andy misses an apple throw and the apple bounces back to the ground, where it started. Similarly, when Andy disappears at the end of an activity, he is present at the start of the next activity as usual. The absent cloud at the start of a growing game is an example of a DDO that affects an object central to the main activity. This DDO is resolved by making the absence temporary. The cloud is programmed to re-enter the screen after a certain delay—just long enough for the child to have noticed its absence and perhaps commented on it. This happens whether or not the child notices or does anything. Even if a child is confused and does not really understand what happened, the cloud returns after a fairly short wait and s/he is able to play as usual.

As a final tool for within-game resolution of "bottlenecks" and errors, the keyboard hot keys allow the researcher to un-stick Andy or objects, should they freeze and block a child's action. This is expected to be less disruptive for a child than needing to re-start a frozen activity, losing his progress.

VARIETY The basic game play (i.e. inherently novel content) and designed DDOs seek to provide a mix of passively observable and actively discoverable elements, through different modalities. The DDOs affect different objects (and Andy), and vary in the extent to which they actively try to attract a child's attention during the games. The goal is always to engage *all* children with *some subset* of DDOs, being aware that what may particularly interest one child may be effectively invisible to another.

Andy's Garden deliberately includes some subtle DDOs that many children may *not* detect, such as removing the fence from the garden background. It also deliberately explores DDOs that are more cognitive, for lack of a better term, than simple changes to appearances and sound effects. Examples of these more "cognitive" DDOs would be those in which Andy makes statements inconsistent with the environment, or those that represent "impossible" events (one apple becomes two, a carrot becomes a strawberry). These inclusions are all opportunities to find out more about the type of aspects and events that children may notice and detect as discrepant.

FREQUENCY The altered games versions should expose each child to a *minimum* of 26 designed DDOs per session, with the expectation that there will be additional opportunities related to genuine system errors, touch-screen use, and perceived inconsistency. This is in relation to a games session estimated to last 15 minutes, or slightly less[13]. This would mean encountering (or being passively exposed to) *at least* 1.6 DDOs every minute. This is near the number of DDOs estimated as a minimum for what designers could include, per actual DR pairs in study 1. Even with non-designed DDOs, the total number that children encounter is estimated to be manageable, when balanced against the substantial sameness within and around *Andy's Garden*.

AMBIGUITY The DDOs in *Andy's Garden* closely follow original discrepant aspects in ECHOES, to which children reacted with a variety of behaviours and for a variety of reasons. It is anticipated that the new DDOs offer similarly open opportunities for communication. They are things to communicate *about*, not demands for specific communication.

[13]This was an initial estimate for the purposes of design, and agreed with the playing time of the younger typically developing testers.

7.8 Summary

This chapter has reported the development of the *Andy's Garden* suite of three mini-games. It is designed to be played in three sessions: one with *basic game versions*, used to establish a child's knowledge and expectations about the game contents, and then two sessions in which different *altered game versions* present the child with DDOs. The main source of novelty is *inherent novelty* in session one, where the child initially encounters digital objects and forms expectations about their behaviours. DDOs that attempt to violate expectations include altered object appearances, sound effects, and timings between events. The character Andy also makes occasional "mistakes" with his actions and utterances. These things are expected to interest children and pose opportunities for them to spontaneously initiate communication for a range of goals (e.g. share information or affect, ask a question, request help), and using any behaviours (e.g. speech, gestures, gaze). One involves sorting apples by colour; two centre on growing flowers or carrots by shaking a magic cloud.

Game development began with a set of initial decisions that constrained the form of the eventual design, and some of its contents. These established that the new designs would build directly on ECHOES activity goals and mechanics, and re-use the Andy character in a similar role. Unlike ECHOES, the new games do not try to directly facilitate social skills practice or turn-taking. The current interest is in children's spontaneous communication within and about the games. The games *do* share and re-interpret ECHOES' garden setting, cause-and-effect play, and exploratory, non-competitive format (originally developed with extensive stakeholder input).

Initial game concepts, based on the ECHOES *Cloud Raining* and *Ball Sorting* activities, were implemented in the GameSalad game engine to create the basic game versions (i.e. those with "normal" behaviour and no DDOs). One game involves sorting apples into a set of tubes, based on their colour (Tidying up the apples). Two extremely similar games have the child use a cloud to grow colourful flowers or carrots in the garden (Flower Growing and Carrot Growing). The basic mechanics of the growing games are identical, though the fully grown plants have different properties and opportunities for interaction. Development was shaped by input from a design critique with academic experts in HCI and children's technologies, focusing on issues of how the VC Andy can meaningfully model actions for children, and use limited dialogue to communicate activity goals. The final basic game versions were copied and modified to create altered game versions with DDOs. Generating, choosing, and balancing DDOs had a large number of constraints at the level of individual mini-games, and games sessions. The goal was to create a set of DDOs in each activity and session that represented high frequency and variety, allowed flexibility and resolvability of actions, and provided ambiguous opportunities for communication. The eventual set represents a range of objects and modalities, and mixes opportunities for children to passively observe versus actively discover DDOs. Final basic and altered games were usability tested with a group of TD children. This yielded overall positive feedback, but also highlighted ongoing challenges with object physics. Without detailed analysis, it appeared that the TD children were noticing and reacting to many of the designed DDOs. Informal, critique-type discussion identified that children strongly desired more interaction and responsiveness from Andy. The basic and altered game versions were updated following this testing, to improve usability and provide more interactivity.

The development process has highlighted that designing "motivating but manageable" games requires consciously designing *sameness* as well as differences, in the form of DDOs. The need to provide engaging, varied, and balanced experiences also requires considering the games session as a larger unit of content, i.e. the amount that a child is likely to play on a single day. The principles detailed in Chapter 6 must be realised collectively in the technology and its manner of use. The *Andy's Garden* suite of games were evaluated in a small-scale school study with new child participants, reported in the following chapter.

Chapter 8

Study 2: Proof-of-concept study with Andy's Garden

8.1 Introduction

The main goal of the current study is to evaluate the success of a new set of games, *Andy's Garden*, in motivating child initiations about situations that children determine are inconsistent with their own expectations, i.e. *discrepancies*. The current games have been designed as a tool to test the final research question in the "line of investigation" laid out at the start of the thesis:

> [RQ5] Guided by design principles rooted in empirical work with ECHOES and autism theory, can designed *discrepancy-detection opportunities* (DDOs) successfully create a *motivating but manageable* experience with new participants, in a new set of games?

Chapter 6 has outlined design guidance for attempting to deliberately create *discrepancy-detection opportunities* (DDOs), and how to include these in a motivating, emotionally manageable way. Chapter 7 described the design and development process for *Andy's Garden*, the set of three new mini-games used in the current study: Flower-Growing, Tidying up the Apples, and Carrot-Growing. This chapter outlines the research questions and methodology for a proof-of-concept scale evaluation study of *Andy's Garden*. It reports three general categories of results: those related to children's communication in and around the games, those that report child reactions in relation to different design elements, and qualitative examples of child interactions. This last give a clearer sense of what child behaviours were like. The discussion reflects on the extent to which *Andy's Garden* has provided proof-of-concept for DDOs as a strategy to motivate initiation, and the success of different design decisions.

8.2 Research questions and study design

8.2.1 Research questions

The current study has a dual purpose: to explore children's social communication around a set of new games, and also how specific design decisions impact children's communication and their overall game experience. Consequently, there are two interlinked sets of research questions, and two interlinked sets of results.

8.2.1.1 Communication-focused research questions

The results of Study 1 raised new questions about whether communication about discrepancy was a general phenomenon, and if this pattern of interaction could be feasibly re-created in another context, with another group of children. The main questions regarding child communication in *Andy's Garden* are as follows:

RQ5-a Will children be motivated to spontaneously initiate social communication about the designed DDOs?

RQ5-b Will children's game-play experience be emotionally manageable, per their affect, utterances, and behaviour during games sessions?

At the highest level, this study tries to determine whether it is possible to deliberately re-create the patterns of positive child interaction that naturally emerged around the ECHOES virtual environment, through DDOs. Re-creating an ECHOES-like pattern of interaction would further mean:

1. Including designed events and situations that children notice, and react to as being discrepant in some way;

2. That all children in a group react to some DDOs, but may differ widely in the things to which they react;

3. That a sizeable portion of all children's reactions to discrepancy (DR pairs) are initiations to a partner;

8.2.1.2 Design-focused research questions

If child initiation about discrepancy appears to be present, to any extent, then it makes sense to ask more detailed questions about the *overall effect* of the design. There is also an aim to explore specific decisions and elements within the design. This fills in the picture of RQ5-a and RQ5-b, above, but also addresses RQ5-c: How will specific design decisions affect children's communication and their overall game-play experience?

More specifically, this means:

1. Which particular designed DDOs and designed novel game elements appear to have been successful—or unsuccessful—in motivating child initiations?

2. How did specific design decisions appear to contribute to, or detract from, children having a "motivating but manageable" game experience?

3. Across all DR pairs, what features[1] seem to be particularly effective (or not) in attracting children's attention, and motivating them to react or initiate?

4. What is the nature of the relationship between the design principles articulated in Chapter 6, and the results observed in this study?

As so little is known about discrepancy as a possible motivator of communication, these questions are exploratory by necessity. They are intended to generate fruitful avenues of investigation for future designs, more so than to make precise judgements or rankings of "success".

8.2.2 Study design

The study used a within-participants design, in which all children played an identical set of games, with the same order of content (described in more detail in Chapter 7). It seeks to understand children's behaviour individually, rather than comparing "performance" between children. The study focuses exclusively on child reactions to discrepancy, and whether and how DDOs can motivate communication. It does not consider reactions to discrepancy versus those to other game elements. It is premature to ask questions of this type: too little is known about this area. This study does not

[1] E.g. modality, relationship to activity goal.

have a control group of children with ASC (playing non-DDO games), nor a comparative group of typically developing child participants, for the reasons discussed in Chapter 3.

8.3 Methodology

8.3.1 Ethics

8.3.1.1 Ethical approval and philosophy

This study received ethical approval from the University of Edinburgh School of Informatics. Permission for this particular study to be carried out in the target school was also granted through the local council's Children and Families research access process, which is similar to a standard university ethical approval process, but with the greatest emphasis on assessing the participation time/tasks required from school staff members, and ensuring confidentiality. All parent and child information and consent forms are reproduced in Appendix H.

This project advocated an approach to consent that attempted to respect both the conviction that children should be offered the opportunity to give informed consent (or refusal) to research on their own behalf, and the concern that young children (especially those with autism) may not understand some types of information necessary to make a fully-informed choice. The current "standard" practice of child assent and parental proxy consent puts the main burden of the consent process on the parent. Child assent/consent is generally not treated as a requirement (for ethics committees, school administration, etc.) in the same way as is parental consent. The project planned to follow a strategy here termed *most-informed consent.* It includes a parent component and a child component in the consent process, but views the researchers' approach to parents as one of the layers of "gate-keeping" required for researchers to access children as potential participants: schools must agree to host the research, teachers must agree to open their classrooms to researchers, and then parents must agree that researchers can approach their child about participation in a specific piece of research, with certain projected risks, benefits, and dissemination of results. Parents indicated their agreement that their child could participate, giving consent for the researcher to approach the child abut participation, rather than giving proxy consent. Children may still elect not to participate, even if parents have given permission. In this project, approval/refusal of some data use options remained with parents (such as whether or not images may be used in publications). Ideally, the child and researcher would than have a short "consent conversation" about taking on a potential role (that of games tester, or helper). This consent is "most-informed consent" because each young child will receive as much information about the proposed research participation as appears feasible for the him or her, in that context and with respect to his or her usual methods of communication, language comprehension, and other needs. The researcher will likely need to consult other adult stakeholders about those issues. Hughes and Helling (1991) maintain that pre-school aged children *can* understand the key ideas of research and the consent process if researchers respect their needs by choosing concrete language and clearly relating the research activities to the child's life experiences and contexts (see also Fraser et al., 2004). The overarching goal is for each child to be "as informed as possible" or "as informed as feasible" in making his or her choice.

8.3.2 Participants

The current analysis includes data from ten children with ASC (M=8, F=2) with a mean developmental age of 5 years, 9 months (range 3:02 -7:05)[2]. Children all attended the same medium-sized, autism-specific special school in Scotland, and were drawn from several different classes. Teachers were asked to identify children in their classes who could feasibly be invited to participate, based on the following inclusion criteria:

- Developmental age of 3-6 years (any chronological age)
- Minimum language ability: can understand and produce short verbal phrases (even if production is infrequent)
- Able to participate in an activity for about 10 minutes, with adult support
- Unlikely to be distressed by working with computers, or working with relatively unfamiliar people
- NO reading or writing ability required

Teachers were further advised to "keep in mind that the priority is not to identify the children who may be most able or 'best' at doing tasks, but those who have enough language to understand the instructions, and will not find the study experience scary or 'unmanageable'". They identified 16 potential participants across the primary classes,

[2]As estimated via the British Picture Vocabulary Scale (BPVS) measure of receptive language, see Section 8.3.2.

who then received informational study materials and consent forms through the school administration. Of those children, 12 received parental consent, and participated in games sessions. Two children were later excluded from the analysis. One child, Hassan, was determined not to have met the initial inclusion criteria regarding his language production. He customarily used an AAC iPad app for the majority of his communication, and had almost no verbal language production at all. The child was not able to use the study computer and his iPad at the same time: having them both present proved far too distracting. He did play the games, but without his iPad available—thus severely limiting his opportunities for expression. The second child, Colin, met all inclusion criteria, but had substantially incomplete data because he did not finish many activities within the game sessions. Given the game designs, which aimed to test the research questions by presenting items and events in a particular order to establish, then try to violate, expectations, this meant that his data could not usefully be part of the analysis because he saw only some parts of the "normal" game versions.

As all children must have a documented ASC diagnosis to attend the school, this project did not seek to re-diagnose children. As in ECHOES, two standard measures were used to estimate children's linguistic and social ability: the Social Communication Questionnaire (SCQ; Rutter et al., 2003), completed by parents, and the British Picture Vocabulary Scale, 2nd edition (BPVS; Dunn et al., 1997), a measure of receptive language ability. BPVS scores were collected by a trained researcher immediately prior to the evaluation, except for one child who had a recent BPVS score (3rd edition; Dunn, L. M., Dunn, D. M., 2009), obtained from the school's speech and language therapist (SLT). Table 8.1 reports participants' demographic information and test scores, including verbal-mental ages (VMA) calculated based on BPVS scores. The disparities between VMA and chronological age suggest that almost all participants had some degree of intellectual disability in addition to ASC. Unexpectedly, one child (Leanne), appears to have a higher VMA than her chronological age.

Prior to conducting the experimental sessions, the researcher spoke to each child's teacher about his or her support needs, and how best to interact with him or her. The researchers also engaged in additional observation and activities to make themselves known to the children, and reduce the anxiety of the first study session (described in more detail in 8.3.3.1). Table 8.2 summarises additional information on each child, which form a helpful supplement to their standardised scores and a useful context in which to view their interaction during the study.

Child (Gender)	Age (year:months)	SCQ	BPVS: Raw	BPVS: VMA
Alfred (M)	5:03	17	54	5:04
Arthur (M)	11:02	32	52	5:01
Bradley (M)	9:09	34	41	3:10
Campbell (M)	6:05	21	54	5:04
Connor (M)	7:07	31	58	5:08
Eshan (M)	7:08	18	31	3:02
Kristina (F)	7:09	16	29	4:01[a]
Leanne (F)	6:07	20	76	7:05
Martin (M)	6:02	31	39[b]	5:03-5:05
Omri (M)	7:05	22	38	5:01[a]

Table 8.1: Participant information and standardised measures for Study 2 (pseudonymous).

a. Calculated via BPVS 2nd edition norms tables for English as an Additional Language (EAL), based on teacher information about child's language background.

b. BPVS 3rd edition raw score and VMA equivalent obtained from school SLT

8.3.3 Materials and procedure

8.3.3.1 Familiarisation, class observations, and teacher consultations

Many children with autism can find it disruptive to have unfamiliar people introduced to their classroom environment, or to take part in non-routine activities. Introducing the researcher to participating children beforehand, often in a whole-class context, has the dual purpose of making the researchers known to the children, and allowing researchers to get a sense of their participants. Class observation is also extremely important for researchers to learn the customary phrases (and/or modified sign language) used in a school for things like indicating the start and end of activities, offering choices, giving praise, or encouraging children to sit, listen, and attend.

For this study, there were four classes with participating children. The researchers observed each class for approximately 1 hour, during normal daily activities. On a subsequent day, the researchers led a drawing-based familiarisation activity in each class, including the non-participating children. This type of activity had proved to be a very successful familiarisation tool in a previous study (e.g. Alcorn, 2010), as children could follow their own interests and work at their own level of ability, while the

Child	Description	Language
Alfred	Teacher notes he often does or says deliberately silly/contrary things to get a reaction. Sensitive to sound, played most of games sessions at a very low volume. Very willing to explore within the game, actively predict how things might work.	Child enjoys talking, and getting others to talk. He has fluent phrase-level to conversational-level language with little or no echolalia. Some gaps in vocabulary. Teacher notes he may sometimes seem "deceptively able" due to his language production.
Arthur	Child was calm, cooperative and engaged in sessions, although teacher notes he often dislikes new things the first time they are encountered. Child previously participated in testing with a very early version of the ECHOES system (2010).	Child can be overwhelmed by too much language or instruction. Infrequently used verbal language but has at least phrase-level ability. Little or no echolalia.
Bradley	Child frequently impassive, seemed dissociated or withdrawn from the school environment, as though "in his own world". During games sessions, frequently distracted by noise and activity outside the room.	Needs simple language, instructions. Child appeared to have at least phrase-level language, but used it infrequently. Teacher confirms he tends to use less language than he knows. Some echolalia.
Campbell	Teacher noted that child needs considerable processing time, struggles with choices, and may be anxious about new things. Difficulty with transitions, things being "finished". Frequently sought reassurance and information from adults.	Child has good phrase-level language, though some difficulty with pronouns.
Connor	Teacher describes child as frequently impulsive, with a short attention span. Must check that he is listening. During games sessions, had low tolerance for pauses, waiting, or touch screen problems.	Produces phrase-level to conversational-level language. Tends to ask many questions, and was especially preoccupied with asking about his classmates and their game performance.
Eshan	Teachers noted a tendency to engage in repetitive/irrelevant questioning. Once he grasped the game mechanics, he became extremely focused. Difficult to persuade to experiment within the game, or wait for character turns.	Produces phrase-level to conversational-level language, but pronunciation/enunciation may be difficult to understand. Finds questions difficult. Regularly and persistently talked about "off-topic" subjects, unclear how much is echolalia.
Kristina	Child seems to passively enjoy social interaction. She is motivated by adult praise and approval and will actively seek it. Teacher describes her as compliant, rarely getting very excited or upset.	Child is a non-native English speaker, who fluently produces phrase-level to conversational-level language. Difficulty with open questions.
Leanne	Child actively approaches her classmates and the researcher, and volunteers information about interests and experiences. Teacher describes as having low anxiety, and being generally "easy to work with".	Child has fluent, conversational-level language. She enjoys talking, and talks frequently and at length.
Martin	Child seems to passively enjoy social interaction, and be motivated by adult approval. Sometimes volunteers information about class activities, experiences.	Child produces fluent verbal language with frequent echolalia. May have less comprehension than suggested by his speech fluency.
Omri	Teachers describe as easygoing, with low anxiety. Prefers to do things alone. Special interest in computers. In games sessions, seemed very rigid in behaviour about right and wrong way to do things, reluctant to explore.	Child is a non-native English speaker, still learning the language. He produces phrase-level to conversational-level language with little echolalia. Pronunciation/enunciation often difficult to understand, some stammering.

Table 8.2: Description of Study 2 participant characteristics.

Figure 8.1: A child drawing from the familiarisation activity, using a blank background sheet. This child has drawn himself and several Hoovers (i.e. vacuum cleaners), a special interest.

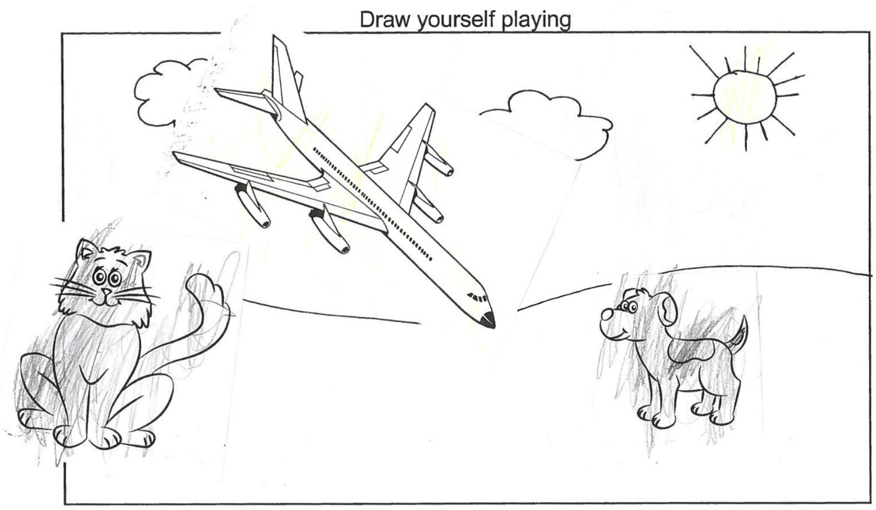


Figure 8.2: A child drawing from the familiarisation activity, using the “outside” background. This child chose several paper cut-outs from a large selection, glued then with the help of a staff member, and has coloured on top.

researchers had an opportunity to attend to and speak to each child in turn. The drawing activity lasted about 20 minutes, and asked children to "draw yourself playing"[3] Additional verbal instructions were required for most children, with the researchers encouraging them to identify at least one specific thing or person they liked to play with. Most children used blank backgrounds (see 8.1), though with classes judged to need more task structure, children began by choosing an "inside" or "outside" background template, in which a minimal line drawing indicated either a room, or a field with sun and clouds. Paper cut-outs[4] and glue sticks were made available for children who had difficulty drawing, i.e. due to motor control issues. Some children used a combination of cut-outs and drawing (see 8.2). No photos or videos were collected during this activity, only copies of the children's drawings (originals were returned to teachers), and researcher notes. The BPVS testing sessions, conducted between class activities and the start of the experimental session, also formed an additional opportunity for familiarisation between the researchers and participating children[5].

As a supplement to the information available through direct observation, the researchers also met with each teacher to discuss the participating children in his or her class. For each child, researchers asked specifically about support needs, sensory sensitivities, challenging behaviours, language, and any other advice about interacting with him/her. A selection of this per-child information has already been reported in 8.3.2. This process also identified a number of general strategies for successfully communicating with the children, rewarding positive behaviour, and minimising any disruptive behaviours. As a result of these discussions, the researchers also provided teachers with pictures of themselves and the games, to be used on class and child visual schedules on days when sessions would take place.

8.3.3.2 Games sessions

Each child individually completed three sessions with the *Andy's Garden* suite of three activities: flower growing, tidying up the apples, and carrot growing. The activities were always presented in this same order during each session, bookended by Andy's hello and goodbye sequences. After playing through each activity once, children were offered the opportunity to "play with Andy some more", and sometimes elected to

[3]This particular choice of prompt was part of a preliminary information-gathering stage for an unrelated project.

[4]These included black-and-white clip-art-style drawings of toys, animals, people, sports equipment, and everyday objects.

[5]All children except Martin, who had a recent BPVS score from the school SLT.

jump to a favourite activity instead of playing through in order again. With some exceptions due to following the school schedule, children were generally able to play as long as they wished on any given day. Individual sessions ranged between 10 and 30 minutes, with an average session length of about 16.5 minutes.

Although originally designed for a 42” multi-touch screen, the games were presented on a 22” touch screen monitor without multi-touch capability. This monitor was a substitute piece of equipment. Please see the discussion for how this may have affected children’s play. Children had the choice of using with touch screen with their finger, or a rubber-tipped stylus. Most children switched back and forth between these options during each session.

Games sessions and pre-study BPVS testing took place in a school meeting room generally used by staff members (see this room set up for games sessions, in Figure 8.3). While quieter than a classroom, some sessions had substantial background noise from the nearby music room, or traffic in the hallway outside. During sessions, children sat within reach of the touch screen with an adult researcher seated next to him or her, with a laptop (see Figure 8.4). The researcher used a graphical user interface in order to set up or re-start sessions, but otherwise the progression through the session was automated, as the order of the activities was set. The researcher sometimes used the laptop keyboard to input additional “hot key” commands, for example for Andy to take a turn or repeat instructions (see Chapter 7). The researcher’s use of the interface was minimal during sessions; her main role was as a social partner. As in the original ECHOES evaluation, researcher-child interactions were not scripted, but guided by the principle of consistently responding to and supporting the child, and minimising direct instructions or modelling of activities. All activity instructions were initially delivered by Andy. Generally, direct researcher instruction or modelling only took place when those were the fastest or clearest ways to prevent a child from becoming frustrated or disengaged, e.g. because s/he could not figure out what to do.

For the research questions in this study, it was particularly important to find out what children *independently* and *spontaneously* investigated or noticed within the game environment. In some instances, the researcher did encourage the child to explore the environment if s/he seemed very focused on some aspects, to the exclusion of other parts of what could be considered “basic” game play (see Section 8.4.2.3). However, a firm principle of researcher interaction was that she would never draw the child’s attention to DDOs, and would as much as possible try not to visibly react to these, unless the child reacted first. Researcher behaviour should not have signalled to the child that something unusual or interesting was occurring.

All children were accompanied by a familiar staff member from their own classroom, who was seated approximately 1m behind the child, off-camera left (Figure 8.4). Staff members were primarily there to provide support if children became upset, and were encouraged to intervene if concerned about the child for any reason. In most children's sessions, staff were effectively observers. In some instances, however, they provided brief advice to the researcher during sessions on how best to give an instruction or offer a choice to a child. In light of previous studies in which some school staff members seemed highly concerned about children finishing tasks or "doing things right", all accompanying staff members were briefed, via a written information sheet (see Appendix I), about the play and discovery purposes of the study. This emphasised the importance of finding out what children could do and discover independently, and that it was not a problem if they did not finish tasks, played slowly or with repetitive behaviours, or stopped to talk and ask questions. This briefing appears to have been successful, and staff members did not express concerns about task completion or "correctness" during this study.

All games sessions were videoed, with child agreement, by a second researcher using a tripod-mounted camera, seated about 1m away from the child. With a few exceptions, children ignored the camera entirely and rarely initiated to the researcher behind the camera.

8.3.4 Video annotation

475 minutes of games session video were annotated using ELAN (version 4.9.1; Max Planck Institute for Psycholinguistics, 2015)[6], with a mean of 47.5 total minutes of video per child (range= 23-73 minutes). All of each child's video data was analysed, with the exception of some short segments of non-analysable video (i.e. system restarts and child rest breaks). Due to scheduling issues, one child, Campbell, completed only the first two games sessions. Unlike Colin (excluded from analysis), he completed all activities within those sessions. He is missing session 3, rather than having incomplete data overall.

This study uses the same annotation scheme as detailed in Chapter 4, and applied in Study 1 (Chapter 5). There are some relatively small modifications, almost all concerned with recording additional design-related information necessary to answer the

[6]For software, please see http://tla.mpi.nl/tools/tla-tools/elan/

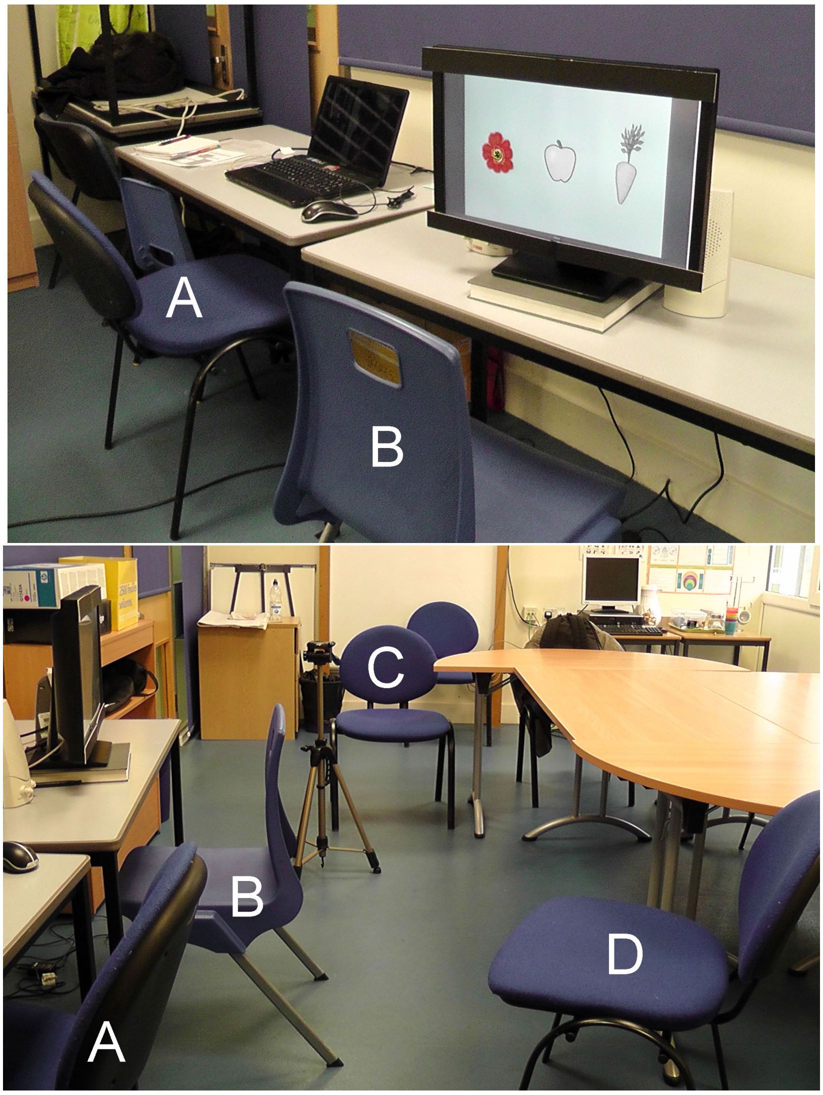


Figure 8.3: Games session set-up in school meeting room, from two viewpoints. Position A indicates researcher, B indicates child, C is the researcher operating the camera, and D is accompanying school staff member.

Figure 8.4: Examples of child and researcher positioning during Study 2.

current research questions (see 8.3.5). However, the annotation definitions and procedures are exactly the same for the 'core' annotation judgements of: for identifying codeable segments, classifying child reactions as social (initiations) or non-social, identifying target partners of initiations, and identifying discrepancy subtypes. The inter-rater reliability of the scheme (on extremely similar Study 1 data) was high, and suggested that while the prevalence of the different categories was very unbalanced, there was very low bias in the coding. Details of second coding and reliability are reported in Chapter 5.

As reported in Chapters 4 and 5, the annotation task consists of two parts: firstly, determining each place where DR pairs occurred in the video, and then, for each pair, determining various features about it: the type of discrepancy (novelty, non-event, or surprise) and type of child reaction. Annotations recorded detailed information about discrepancy-reaction pairs and relevant social partner actions, including the context of each discrepant aspect, the child's low-level reaction behaviours (e.g. pointing, commenting) and the reaction type (primary initiation, secondary initiation, or non-social reaction). The annotation procedure followed in this study was almost identical to that in Study 1 (see Chapter 4). The changes made to the annotation scheme are all related to answering research questions about the game design, rather than those about child initiation. These changes are explained in 8.3.5.

It is important to note that, while the current games have included deliberately-designed situations that children may *likely* experience as discrepancies (i.e. *designed DDOs*), the annotation records all segments that appear to qualify as DR pairs, whether these include a designed DDO or not. The distinction between a discrepancy rooted in the design versus one due to a system error (or other cause) is only meaningful to the researcher, not with respect to the child's experience of the game. All discrepancies are a part of his/her experience, and children's reactions to them are still highly relevant to the broader research questions of whether and how discrepancies can motivate communication, as well as design questions about the nature of a "motivating but manageable" experience. Non-designed-DDO discrepancies can also contribute to the picture of what interests children in the game environment, how they understand things, and what "works best" to engage them.

Addition or change	Description
Annotate all instances of designed DDOs and inherently novel elements	Not all children were exposed to the same DDOs, the same number of times, as some were only discoverable through child exploration. It is important to know not only which opportunities each child took up, but which were (or were not) available to them.
Annotate all instances of objective system error	System errors that motivated reactions may suggest further DDOs, and also give a sense of which errors children find disruptive. Allows better estimation of how much "sameness" versus difference is present in a game session.
Explicitly record Evidence of Expectation Formation (EEF)	Adds an annotation label intermediate between a DR pair and DDO-only. Records instances of a child noticing, forming expectations about an aspect, but where behaviour appears short of a reaction. Formalises the process of counting children's exposure to objects, behaviours, etc. that was implicitly present in the original annotation process (Chapter 4).
Record the underlying source of DR pairs	Identifying which discrepancies were related to deliberate design choices, versus system errors (etc). This is important for assessing the success of different parts of the design, with respect to motivating reactions.

Table 8.3: Summary of changes to annotation categories and labelling, from the annotation scheme reported in Chapter 4.

8.3.5 Additions and changes to video annotation

In order to answer current research questions about the success of various design decisions at attracting children's attention and motivating their communication, it was necessary to collect additional information (i.e. additional to what was collected in study 1, Chapter 5) about what happened during game play, in addition to annotating DR pairs. Table 8.3 summarises the additional types of information collected, which are then described in more detail in the following subsections.

There were additional changes to the annotation procedure made possible by the fact that the current analysis is more focused, with a clearer idea of what information will be needed than was the case in the completely-exploratory study 1. Several types of information derived from free-text descriptions during the spreadsheet analysis stage in study 1 (e.g. the emotional valence of children's reactions, labels for discrepant aspects and errors) were now labelled directly, during the annotation stage.

8.3.5.1 Presence of DDOs, inherently novel elements, and system errors

A DDO-only annotation indicates that an aspect has occurred or is present during the game, and went apparently unnoticed by the child. Indeed the child may in some cases be *unable* to notice it, for example if s/he is facing away from the screen or talking to the researcher. Recording the presence of every designed DDO (whether or not the child appeared to notice or react to it) was necessary for two reasons: to have specific evidence about the designed DDOs to which children had an *opportunity* to react, but did not, and also to count the instances of those designed DDOs that the child needed to actively discover in some way (i.e. that would not automatically be seen a set number of times during game play). If a certain designed DDO rarely motivated reactions because it was rarely present, this might lead to a very different judgement of its success than if we had no information about its presence, and thus, about children's opportunities to react. The same unevenness of opportunity is true of some inherent aspects or properties of the game design, i.e. opportunities for the child to notice and react to novelty. While children automatically encountered most novel aspects and properties in the course of ordinary game play, others needed to be actively discovered (see 8.4.2.3). Thus, a DDO-only annotation was also used to indicate objectively *novel* aspects and properties present in the game, but to which children did not react.

It must be emphasised that DDO-only annotations are not "missed opportunities" for the child to react to discrepancy, because a discrepancy can only be present when the child detects an inconsistency between his expectations and the state of the environment. Rather, they are missed opportunities on the design side, instances in which the design failed to attract the child's attention and motivate him or her. A DDO-only annotation records the state of the environment, only.

The video annotation also recorded the nature and duration of all objective system errors, either as part of DR pairs or as error-only annotations. Keeping track of errors *without* reactions facilitates a clearer understanding of what errors actually disrupt children's game play, versus what could be ignored (or is not even noticed). These errors are part of the overall sameness-difference balance of the game.

8.3.5.2 Evidence of expectation formation

This annotation label essentially formalises the process of counting a particular child's exposure to objects, behaviours, and events in order to determine when s/he may have formed expectations about them, meaning that they are no longer novel and that the child would be able to notice alterations (i.e. surprises). This counting was implicitly present in the version of the annotation scheme used in Study 1, in the "rule of 3" used to determine when a child should be considered to have formed an expectation about a kind. However, it was up to the annotator to keep track of these exposures. The current annotation procedure counts them explicitly, and also now distinguishes between instances in which there is evidence that the child attended to or was aware of something versus when that same element is merely present, with no evidence (DDO-only annotation).

Evidence of expectation formation (EEF) is an annotation label intermediate between a child reaction, and a DDO-only annotation. It is intended to capture instances where the child's behaviour suggests that s/he *notices* or is aware of an aspect, and is presumed to be forming expectations about it, but the child is not motivated to initiate or otherwise alter his/her behaviour enough that it constitutes a reaction. An example from the current dataset is the designed DDO "butterfly flies across screen". Multiple children could be seen to visually track the butterfly, but without commenting, touching it, or, pausing in carrying out the activity. We have evidence that they are aware of the butterfly and gaining some kind of information about it. Thus, we have *evidence of expectation formation* but would not note a DR pair.

The following is a definition of evidence of expectation formation (and its converse) in terms of aspects and kinds, as introduced in Chapter 4:

> **Evidence of expectation formation (EEF):** When a child attends to or interacts with a previously unseen (i.e. objectively novel) aspect of the environment, k, in such a manner as to suggest that s/he is aware of its presence and is gaining knowledge of the properties of aspect k, as a representative of kind K. The knowledge does not need to be detailed, accurate, or complete: the child may have knowledge of only some properties, or may have misunderstood those properties. As a heuristic, after three instances of EEF related to the same kind K, that kind and its child instances should be considered familiar (i.e. non-novel) to that child, unless the child's behaviour provides contrary evidence.
>
> **No evidence of expectation formation:** When a previously unseen (i.e. objectively novel) aspect of the environment, k, occurs or is present but the child does not attend to it or interact with it in such a way as to suggest that s/he is aware of its presence and or is gaining knowledge of its properties.[7]

Note that EEF is not *in itself* a type of child reaction to a discrepancy. It is a

[7] In some instances where the aspect is continuously present throughout an activity (e.g. snail) only one note is needed per activity.

way of keeping track of whether something is able to be counted as novel (i.e. no expectations yet), or if children have the expectations that would allow them to be surprised. There must be at least one piece of evidence of expectation formation in relation to a kind K in order for a child to detect a surprise involving that aspect. If a child's behaviour in relation to a novel aspect qualifies as a reaction (social or non-social), that automatically means that s/he is aware of the aspect and is learning about it; it would be redundant to also annotate EEF.

8.3.5.3 Recording the source of DR pairs

The current annotation added an additional level of labelling to identify the apparent "source" of each discrepancy-reaction pair. This label answers the question: If the child finds a situation discrepant, what was the underlying cause of that situation? Source labels derive from the researcher's knowledge of the system, rather than from the child's behaviour and experience, and are most relevant to the design-related research questions. There are five sources in the current data set:

1. Designed DDOs
2. Inherently novel game elements or properties
3. System errors
4. Child's own perception and expectations (Perceived DDO)
5. Multiple (pattern or sequence of events with mixed sources)

The final 'multiple' source emerged in the course of the annotation when children reacted to patterns of mixed character errors and designed mistakes, and is in some respects a special case of Perceived DDOs. Where a child's perception of discrepancy does not appear to be related to any objective newness, difference, or change in the environment, it is known as a *perceived discrepancy*. In other words, these discrepancies have their source solely in the child's own perceptions or expectations. As an example, the cloud in the growing game was programmed to stop producing rain after all flowers or carrots had been grown. Multiple children expected that the cloud should keep raining after this point, and perceived it as "not working" or being "broken" when it did not rain, commenting to this effect or even asking for help from the researcher. Here, the child's expectations are inconsistent with the world, but the "source" is within the child, not as the result of an objective change, error, or design decision.

8.3.6 Analyses of annotation data

Completed annotations were exported from ELAN as tab-delimited text and further analysed in the Microsoft Excel spreadsheet program. At this point, annotations were double-checked for consistency of labelling within and across children. Each discrepant aspect in the final data set was labelled as being active or passive, depending on the child's involvement in that discrepant situation (for a fuller discussion, see Chapter 6). Excel pivot tables were used to generate summary numbers for communication and design results.

8.4 Results

The focus of the current study was, firstly, to determine whether children would identify and react to discrepant situations while playing the Andy's Garden touch-screen games. Of these reactions, we were interested in determining the relative proportion of non-social to social reactions (where social reactions are initiations to a social partner) and, in the case of social reactions, the target of such initiations (human partner or virtual character). We were also interested in the discrepancy subtypes to which children reacted (novelty, [surprising] non-events, and surprise). The small sample size and within-child observations necessitate non-parametric statistical tests, in this case Wilcoxon signed-ranks. All reported p values are two-tailed, with alpha values set at 0.05.

8.4.1 Child communication results

8.4.1.1 Frequency, target and type of child reactions to discrepancy

Table 8.4 gives counts of each child's D-R pairs, divided by the type of reaction and the subtype of discrepancy which they appeared to be about. Overall, there were 650 child reactions to discrepancy (range= 29 to 121, mean= 65.0, SD=30.02). Of the total child reactions, 409 (62.92%) were directed to a social partner (i.e. an initiation). Table 8.5 gives the count of child initiations (primary and secondary) by target partner. Significantly more of the initiations were directed to the researcher[8] (90.46%) than directed to Andy (9.54%) (T= 0, z= 2.78, p=0.005).

[8]A small number of initiations were directed to other adults present, rather than the researcher primarily interacting with the child. They are included in this category.

Child	Novelty			[Surprising] Non-event			Surprising event			Total
	Non-social reaction	Primary initiation	Secondary initiation	Non-social reaction	Primary initiation	Secondary initiation	Non-social reaction	Primary initiation	Secondary initiation	
Alfred	12	17	4	8	14	2	14	23	5	**99**
	(12.1%)	(17.2%)	(4.0%)	(8.1%)	(14.1%)	(2.0%)	(14.1%)	(23.2%)	(5.1%)	
Arthur	16	11	1	9	5		10	8	0	**60**
	(26.7%)	(18.3%)	(1.7%)	(15.0%)	(8.3%)	(0.0%)	(16.7%)	(13.3%)	(0.0%)	
Bradley	5	1	0	4	3		9	6	1	**29**
	(17.2%)	(3.4%)	(0.0%)	(13.8%)	(10.3%)	(0.0%)	(31.0%)	(20.7%)	(3.4%)	
Campbell	5	6	1	3	13	5	7	2	0	**42**
	(11.9%)	(14.3%)	(2.4%)	(7.1%)	(31.0%)	(11.9%)	(16.7%)	(4.8%)	(0.0%)	
Connor	4	4	0	3	16	2	2	4	2	**37**
	(10.8%)	(10.8%)	(0.0%)	(8.1%)	(43.2%)	(5.4%)	(5.4%)	(10.8%)	(5.4%)	
Eshan	5	6	2	12	10	5	1	1	0	**42**
	(11.9%)	(14.3%)	(4.8%)	(28.6%)	(23.8%)	(11.9%)	(2.4%)	(2.4%)	(0.0%)	
Kristina	15	15	2	14	22	9	13	22	9	**121**
	(12.4%)	(12.4%)	(1.7%)	(11.6%)	(18.2%)	(7.4%)	(10.7%)	(18.2%)	(7.4%)	
Leanne	6	7	1	7	8	2	4	16	5	**56**
	(10.7%)	(12.5%)	(1.8%)	(12.5%)	(14.3%)	(3.6%)	(7.1%)	(28.6%)	(8.9%)	
Martin	6	7	0	10	15	5	16	12	11	**82**
	(7.3%)	(8.5%)	(0.0%)	(12.2%)	(18.3%)	(6.1%)	(19.5%)	(14.6%)	(13.4%)	
Omri	2	8	4	11	20	7	8	12	10	**82**
	(2.4%)	(9.8%)	(4.9%)	(13.4%)	(24.4%)	(8.5%)	(9.8%)	(14.6%)	(12.2%)	
Total	**76**	**82**	**15**	**81**	**126**	**37**	**84**	**106**	**43**	**650**
	11.7%	12.6%	2.3%	12.5%	19.4%	5.7%	12.9%	16.3%	6.6%	
Mean	7.6	8.2	1.5	8.1	12.6	4.6	8.4	10.6	4.3	65.0
	12.4%	12.2%	2.1%	13.0%	20.6%	5.7%	13.3%	15.1%	5.6%	
SD	4.9	4.9	1.5	3.8	6.1	2.6	5.1	7.8	4.4	30.0
	6.3%	4.3%	1.9%	6.1%	10.5%	4.4%	8.3%	8.0%	4.9%	

Table 8.4: All child reactions to discrepancy in Study 2, by reaction type and discrepancy subtype.

Child	Initiation to human	Initiation to VC
Alfred	53 (81.54%)	12 (18.46%)
Arthur	24 (96.00%)	1 (4.00%)
Bradley	9 (81.82%)	2 (18.18%)
Campbell	26 (96.30%)	1 (3.70%)
Connor	23 (82.14%)	5 (17.86%)
Eshan	23 (95.83%)	1 (4.17%)
Kristina	75 (94.94%)	4 (5.06%)
Leanne	36 (92.31%)	3 (7.69%)
Martin	46 (92.00%)	4 (8.00%)
Omri	55 (90.16%)	6 (9.84%)
Total	370 (90.46%)	39 (9.54%)
Mean	37 (90.30%)	4 (9.70%)
SD	19.92 (6.17%)	3.35 (6.17%)

Table 8.5: Child initiations to the researcher and virtual character, as counts and as percentages of a child's total initiations.

Child reactions are divided fairly equally across the three discrepancy subtypes: surprising events (35.85%), [surprising] non-events (37.54%), and novelty (26.62%). The non-events were dominated by instances in which a child had difficulty with the touch-screen interface, making his/her actions ineffective, or objects "not work", accounting for 131 out of 241 non-event DR pairs (53.69%), or 20.15% of total DR pairs. Children frequently commented and requested researcher help with the interface, sometimes resulting in prolonged sequences of cooperative interaction.

In this dataset, there were 526 unique discrepant aspects to which children reacted (range= 26 to 94 per child, mean= 55.5, SD=24.49). This number is lower than the total child reactions to discrepancy because it excludes secondary initiations.[9] By definition, secondary initiations are in addition to a primary initiation about a given aspect that the child considers discrepant, i.e. they are additional counts in relation to the same (unique) discrepant aspect. Of these 526 unique D-R pairs, 232 (44.19%) included non-social reactions while 294 (55.80%) included primary initiations to the adult or to Andy. This difference between the number of primary initiations and non-social

[9]It also excludes 29 instances of "duplicate" reactions in which the child made both a non-social reaction and a primary initiation, or two distinct non-social reactions, in relation to the same discrepant aspect. Only the *first* reaction to each unique aspect is included in the total.

reactions was not significant (T=10.5, Z=1.72, p=0.086). Considering only unique discrepant aspects, the balance between the three discrepancy subtypes remains similar, with 31.75% surprising events, 38.21% non-events, and 30.04% novelty.

Of the 294 primary initiations, 76 (25.85%) were followed by at least one secondary initiation, constituting a sequence of child initiations. There were 95 secondary initiations in total (range=1-21, mean= 9.5 per child, SD= 6.48), though the percentage of each child's primary initiations that continued on to secondary (and subsequent) initiations varied considerably, from 4.17% to 42.11% (mean= 24.14% continuing, SD=12.29%).

8.4.1.2 Rate of reaction to discrepancies

Children varied considerably in the length of their game play, and thus their opportunities to communicate. Considering rates of reaction per minute of video allows us to estimate how much each child reacted, relative to his/her amount of video. It is important to keep in mind that, while these average rates are useful, they do not give a good sense of the rhythm of child-partner interaction within a session; this was often uneven, with periods of concentrated interaction, and then periods of the child being absorbed in the game. Table 8.6 reports the rate at which each child reacted to discrepant aspects. The mean rate across all children was 1.4 reactions per minute (range 0.49-1.89 reactions, SD= 0.41), or a mean of 4.24 for every 3 minutes of play. Considering the mean rate of initiations only (Table 8.7), children made 0.89 per minute (range 0.18-1.38 initiations, SD= 0.30), or 2.6 initiations for every 3 minutes of play. Bradley appears to be an outlier here, and is by far the least responsive; the next-least-responsive children react and initiate more than twice as often as he does.

8.4.1.3 Emotional valence of child reactions to discrepant events

We looked at the emotional valence of children's reactions to discrepancy, to see if the earlier pattern of positive reactions was also present in this new data set. The vast majority of child reactions appeared to have a positive (29.69%) or a neutral (59.23%) valence, as judged by children's verbalisations, facial expressions, gestures, and movement. Children varied widely in the degree to which they expressed affect, with some showing fairly "flat" affect throughout their sessions, with rare but visible instances of excitement or pleasure suggesting they were highly engaged by a particular discrepancy. Instances of clearly identifiable but mild negative affect (as defined in 5) were

Child	Total reactions	Video (min)	Reactions per minute
Alfred	99	73	1.36
Arthur	60	47	1.28
Bradley	29	59	0.49
Campbell	42	23	1.83
Connor	37	29	1.45
Eshan	42	30	1.40
Kristina	121	64	1.89
Leanne	56	44	1.27
Martin	82	62	1.32
Omri	82	44	1.86
Total	650	475 minutes	1.37 reactions/minute
Mean	65	47.5 minutes	1.41 reactions/minute
SD	30.02	16.79 minutes	0.41 reactions/minute

Table 8.6: Rates of child reaction to discrepancy, per minute of annotated video. Includes duplicate reactions.

Child	Initiations	Video (min)	Initiations per minute
Alfred	65	73	0.89
Arthur	25	47	0.53
Bradley	11	59	0.19
Campbell	27	23	1.17
Connor	28	29	0.97
Eshan	24	30	0.80
Kristina	79	64	1.23
Leanne	39	44	0.89
Martin	50	62	0.81
Omri	61	44	1.39
Total	409	475 minutes	0.86 initiations/minute
Mean	40.90	47.5 minutes	0.89 initiations/minute
SD	21.90	16.79 minutes	0.35 initiations/minute

Table 8.7: Rates of child initiation to discrepancy, per minute of annotated video.

Child	Positive	Neutral	Negative	Unclear	Grand Total
Alfred	26 (26.26%)	66 (66.67%)	5 (5.05%)	2 (2.02%)	99
Arthur	30 (50.00%)	26 (43.33%)	3 (5.00%)	1 (1.67%)	60
Bradley	10 (34.48%)	17 (58.62%)	2 (6.90%)	-	29
Campbell	9 (21.43%)	29 (69.05%)	4 (9.52%)	-	42
Connor	1 (2.70%)	24 (64.86%)	12 (32.43%)	-	37
Eshan	11 (26.19%)	27 (64.29%)	3 (7.14%)	1 (2.38%)	42
Kristina	43 (35.54%)	68 (56.20%)	9 (7.44%)	1 (0.83%)	121
Leanne	29 (51.79%)	24 (42.86%)	2 (3.57%)	1 (1.79%)	56
Martin	12 (14.63%)	54 (65.85%)	16 (19.51%)	-	82
Omri	22 (26.83%)	50 (60.98%)	10 (12.20%)	-	82
Total	193 (29.69%)	385 (59.23%)	66 (10.15%)	6 (0.92%)	650
Mean	19.3 (29.69%)	38.5 (59.23%)	6.6 (10.15%)	1.2 (0.92%)	65
SD	12.79 (14.93%)	19.03 (9.34%)	4.86 (8.85%)	0.44 (0.99%)	30.02

Table 8.8: Total child reactions by estimated emotional valence

relatively uncommon in the data set, comprising 66 out of 650 reactions, or 10.15% of the total. There were zero examples of children becoming severely dysregulated during games sessions, either in apparent connection to discrepancy or for any other reason (i.e. no challenging behaviour, obvious distress, or physical withdrawal from the study space). Some children demonstrated more frustration and discouragement than others, with a range of 3.57% to 32.43% negative reactions (mean 10.88% per child, SD= 8.86%), see Table 8.8. The vast majority of the negative reactions were related to difficulty with the touch screen (45 out of 66, 68.18%). The remainder appeared as 1-3 reactions each to a wide range of aspects that, in some way, represent the game "not working" or blocking the child from reaching his or her goals.

8.4.2 Design results

8.4.2.1 Discrepancy sources

From a designer's—rather than a child's—viewpoint, each discrepant aspect may have one of five sources: it is either related to a system error, an inherent feature or property of the game environment, a designed DDO, or, alternately, may originate in the child's perception only. Close to a quarter of the unique discrepant aspects (24.14%) were designed DDOs, and a further quarter (26.43%) were related to features and properties of the environment that were inherently novel to the child. System errors were responsible for the largest proportion of DR pairs (34.41%). Of these, by far the most prevalent type of error was difficulty with manipulating objects on the touch screen interface (see also 8.4.1.1). A further 14.45% of DR pairs were "perceived discrepancies", or those in which the child's expectations about the game (or the world) were violated in some way, but *not* as the result of a designed DDO or system error. An example would be a child expressing his surprise and disappointment that the apples could not be re-attached to the tree, or initiating to the researcher during a long pause because she thinks something should be happening now, but is not. A small number of DR pairs (0.57%) had multiple sources, when the child reacted to and talked about Andy having made a sequence of mistakes which included both designed errors, and genuine system errors. Only one child, Omri, reacted to discrepancies of this type.

Another important, broad distinction about types of DR pairs—that crosses discrepancy subtypes and the sources mentioned above—is the child's involvement in a discrepant situation. Was the perceived discrepant aspect something the child passively heard or observed, or something that s/he actively discovered through direct interaction?[10]. Here, 54.94% of the unique discrepant aspects were active, 40.68% were passive, and 4.37% were to aspects that blended passive and active components ('both'). Table 8.9 lists the number of active and passive discrepancies to which each child reacted.

8.4.2.2 Designed DDOs

Children could be exposed to a maximum of 31 different designed DDOs in the current game. Not all children had the opportunity to react to everything, because some designed DDOs needed to be discovered through child exploration. Table 8.10 reports these aspects, divided by child involvement. For each aspect, the table lists the ses-

[10]See Section 6.2.2 for more on the distinction between passive and active discrepancies.

Child	UD Active	UD Passive	UD Both
Alfred	48 (48.48%)	41 (41.41%)	10 (10.10%)
Arthur	32 (53.33%)	23 (38.33%)	5 (8.33%)
Bradley	9 (31.03%)	15 (51.72%)	5 (17.24%)
Campbell	32 (76.19%)	10 (23.81%)	-
Connor	22 (59.46%)	14 (37.84%)	1 (2.70%)
Eshan	26 (61.90%)	12 (28.57%)	4 (9.52%)
Kristina	78 (64.46%)	40 (33.06%)	3 (2.48%)
Leanne	26 (46.43%)	28 (50.00%)	2 (3.57%)
Martin	49 (59.76%)	21 (25.61%)	12 (14.63%)
Omri	46 (56.10%)	27 (32.93%)	9 (10.98%)
Total	368 (56.62%)	231 (35.54%)	51 (7.85%)
Mean	36.80 (55.72%)	23.10 (36.33%)	5.67 (7.96%)
SD	19.23 (12.10%)	11.02 (9.49%)	3.81 (5.65%)

Table 8.9: Count and percentage of child reactions, by child involvement in the discrepancy

sion in which it appeared, the number of children who potentially and actually were exposed to it, those who reacted (DR pairs) and those who showed *evidence of expectation formation* (EEF, see 8.3.5.2) about it. This table also reports the number of different children reacting to each designed DDO. Subtotals (e.g. for the Active category) thus list total reactions, with the number of different children who reacted in a given way (primary initiation, non-social reaction) to anything in that category, up to the total of 10 participating children.

It is important to note that, by default, EEF was not recorded for sound-only DDOs (e.g. Andy new feedback, flower fanfare changed to bike horn). With rare exceptions, it was too difficult to judge EEF about these DDOs. In the absence of a definite child *reaction* to a sound DDO, these were annotated as being DDO-only instances. For consistency, the annotation process erred on the side of conservatism.

8.4.2.3 Inherent features and properties

Initially discovering the digital environment and game mechanics provides many opportunities for children to react and initiate. Unlike designed DDOs, it is not possible to give an exact estimate of how many inherently novel features were in the environment–

Designed DDO (Session)	PRI (Children)	NSR (Children)	Total reactions (Children)	EEF (Children)	DDO-only (Children)	Total Instances (Total children exposed)
Active						
Carrot turns into strawberry when picked (3)	9 (6)	1 (1)	10 (6)	13 (8)	-	23 (9)
Flowers: ALL diff (3)	4 (4)	4 (3)	8 (5)	19 (8)	-	27 (9)
Flowers: ONE diff (2)	3 (3)	4 (3)	7 (5)	7 (5)	-	14 (10)
Giant rain (2,3)	-	3 (2)	3 (2)	-	36 (9)	39 (9)
Some carrots giant-sized when picked (2)	7 (3)	8 (5)	15 (5)	12 (5)	-	27 (9)
Subtotal Active DDOs	23 (8)	20 (7)	43 (9)	51 (10)	36 (9)	130
Passive						
Andy apple mistake: throw to wrong tube (2,3)	3 (1)	3 (3)	6 (3)	2 (2)	2 (2)	10 (6)
Andy apple mistake: throws and misses tubes (3)	1 (1)	4 (3)	5 (4)	4 (2)	-	9 (4)
Andy change end-of-activity utterance (2,3)	2 (1)	1 (1)	3 (1)	3 (3)	24 (9)	30 (10)
Andy disappears at end of activity (3)	9 (5)	1 (1)	10 (6)	3 (3)	4 (3)	17 (9)
Andy mis-statement (2,3)	10 (3)	2 (2)	12 (4)	-	35 (9)	47 (10)
Andy new feedback: You did it! (2)	-	-	-	-	19 (10)	19 (10)
Butterfly flies across screen (2,3)	3 (2)	1 (1)	4 (3)	15 (8)	10 (6)	29 (10)
Changed growing behaviour: Partly grown at start (2)	-	-	-	4 (4)	11 (9)	15 (10)
Changed growing behaviour: Plant grows in 1 rain (3)	-	1 (1)	1 (1)	2 (2)	23 (9)	26 (9)
Cloud sound changed: Thunder (3)	-	1 (1)	1 (1)	-	21 (8)	22 (8)
Cloud sound missing (2)	-	-	-	-	25 (10)	25 (10)
Fanfare changed: Bike horn (2)	-	3 (3)	3 (3)	2 (2)	10 (8)	15 (10)
Fanfare changed: Chimes (3)	-	-	-	1 (1)	15 (9)	16 (9)
Fence gone (3)	-	-	-	-	16 (9)	16 (9)
Fireworks reward sound changed: magic noise (2)	-	2 (2)	2 (2)	1 (1)	9 (7)	12 (9)
Planter colour changed (3)	-	-	-	1 (1)	14 (9)	15 (9)
Rainbow appears at activity end (3)	-	-	-	11 (6)	4 (3)	15 (8)
Some apples giant-sized (2)	1 (1)	-	1 (1)	12 (7)	3 (2)	16 (10)
Tube sound changed (3)	-	-		1 (1)	25 (9)	26 (9)
Tube sound missing (3)	3 (2)	6 (4)	9 (6)	5 (3)	16 (4)	30 (9)
White tube produces 2 apples (2)	1 (1)	1 (1)	2 (1)	2 (2)	7 (4)	11 (6)
White tube produces strawberry (3)	5	6 (3)	11 (6)	3 (2)	-	14 (8)
Subtotal Passive DDOs	38 (9)	32 (9)	70 (9)	72 (10)	293 (10)	435
Both						
Cloud missing at activity start (2,3)	8 (6)	5 (4)	13 (6)	9 (4)	9 (6)	31 (10)
Snail added to garden (2)	3 (3)	2 (2)	5 (4)	16 (9)	6 (4)	27 (10)
Some apples do not fall from tree (3)	2 (1)	-	2 (1)	18 (8)	-	20 (9)
Subtotal Both DDOs	5 (7)	2 (5)	7 (8)	34 (9)	6 (8)	47
Total Designed DDOs	66 (10)	54 (10)	120 (10)	157 (10)	335 (10)	612

Table 8.10: Child reactions, expectation-formation, and exposure to designed DDOs

the answer changes depending on the level of detail at which one looks, which actions or effects are considered as “units” or sequences versus in terms of their separate steps, component modalities, and so forth.

Table 8.11 reports child reactions to the elements and properties that children *must* encounter in the ordinary course of game play, in order to be doing the activities at all. This includes knowledge of the garden settings and their contents, basic child actions (e.g. using the cloud to make rain) and character feedback. The design deliberately includes a high degree of repetition both within and across activities. Even playing each activity version through once, children will have seen each of these basic elements enough times such that they are assumed to have noticed these elements and formed expectations about them to some degree. For example, the character Andy’s introduction sequences that play at the start of *every* mini-game, in which he says the same instruction about the goal (i.e. “Look, a cloud! Let’s grow some flowers.”). As exposure has definitely occurred and expectation formation is assumed, Table 8.11 reports the 26 elements of this type to which there were child reactions. As with the results for designed DDOs, this table also reports the number of different children reacting to each element. Subtotals (e.g. for the Active category) thus list total reactions, with the number of different children who reacted to anything in that category, up to the total of 10 participating children.

Other inherently novel elements and properties were present from session 1, but children would not necessarily encounter or learn about them in the course of basic game play. They would only be seen or discovered if a child took certain actions, or, sometimes, was able to pause, wait, and watch the virtual character. For example, a child can play the apple game without ever trying the white “magic” tube which changes an apple’s colour. Indeed, several children refused to try this tube, even when prompted! Conversely, children needed to be able to self-inhibit and wait in order to see Andy take a turn at an activity: if they began taking a turn in the middle of his action sequence, Andy would switch to giving feedback instead. Some children only ever saw Andy take his turn once, in their first session. Table 8.12 lists child reactions, expectation formation, and exposure to the 16 inherent elements and properties that a child may discover or encounter in addition to his/her ordinary game play.

While sound-only designed DDOs were by default annotated as DDO-only in the absence of a child reaction (i.e. EEF category used only in exceptional circumstances), sound-only inherently novel elements are slightly different. Here the EEF category was used, because the sound-related elements were an immediate result of a direct

Novel element	**PRI (Children)**	**NSR (Children)**	**Total reactions (Total children reacting)**
Active			
Child discovers menu button properties	-	2 (1)	2 (1)
Child grows purple flower	3 (1)	5 (3)	8 (3)
Child grows red flower	1 (1)	5 (4)	6 (4)
Child makes rain with cloud	-	3 (3)	3 (3)
Child sorts apple	-	3 (3)	3 (3)
Passive			
Andy goodbye sequence (session end)	2 (2)		2 (2)
Andy hello sequence (activity start)	1 (1)	2 (2)	3 (3)
Andy says apples end sequence	2 (2)	2 (2)	4 (3)
Andy says apples start sequence	1 (1)	-	1 (1)
Andy says carrot start sequence	1 (1)	-	1 (1)
Andy says finished	1 (1)	-	1 (1)
Andy says flowers end sequence	3 (2)	-	3 (2)
Andy says flowers start sequence	1 (1)	1 (1)	2 (2)
Andy says good job	-	1 (1)	1 (1)
Andy says my turn	1 (1)	-	1 (1)
Apple appears in basket	1 (1)	1 (1)	2 (1)
Apples fall off tree at activity start	1 (1)	-	1 (1)
Child reacts to meeting Andy	2 (2)	-	2 (2)
Child reacts to planter boxes	1 (1)	-	1 (1)
Child reacts to seeing apples	1 (1)	-	1 (1)
Child reacts to yellow tube	1 (1)	-	1 (1)
Fireworks reward	7 (5)	5 (5)	12 (7)
Flower fanfare regular	-	1 (1)	1 (1)
Both			
Activity setting: Apple	-	1 (1)	1 (1)
Activity setting: Growing	1 (1)	1 (1)	2 (2)
End of activity; transition	4 (2)	1 (1)	5 (3)
Total	**36**	**34**	**70**

Table 8.11: Child reactions to inherently novel elements required for basic game play

child action (e.g. Andy says hello when child touches his face) or might cause a child action that would show attention and understanding (e.g. Andy asks child to pick a carrot).

8.4.2.4 System errors

There were 191 DR pairs related to system or interface errors. For example, two different animation errors related to Andy's turn-taking appear in Figure 8.5. As noted in a previous section, the majority of system errors to which children reacted had to do with the touch-screen (further discussed in 8.5.3.3). However, many others involved (mis)behaviour of the game contents, and some were actually quite successful in capturing children's interest and motivating their communication. For example, a physics issue in the game engine meant that if apples were dropped or released in a certain way, they would bounce rapidly between the floor and ceiling. Multiple children enjoyed trying to catch them, or even just watching them. Some errors, especially related to the virtual character making mistakes, were extremely similar to designed DDOs. Table 8.13 lists the 30 different types of system errors present in the annotated video, divided by their interactivity, with the number of children who reacted to each. No information on expectation-formation was collected for this category.

8.4.2.5 Perceived discrepancies

Where a child's perception of discrepancy does not appear related to any objective newness, difference, or change in the environment, it is known as a *perceived discrepancy*. In other words, it does not appear to have its source in the inherent novelty of the environment, a design decision, or a system error. Table 8.14 lists the perceived discrepancies in the data set, divided by their level of child involvement (activity or passivity). While sometimes more than one child seemed to share the same perceived discrepancy (e.g. that the researcher had made a mis-statement or mistake), it does not make sense to report EEF or presence-only data for these situations.

8.4.3 Qualitative interaction examples

This section presents a very small selection of child interactions with DDOs and inherently novel elements. The examples are chosen to convey some of this variety of child-system and child-researcher interaction, with a focus on initiation. In one case,

Novel element	PRI (Children)	NSR (Children)	Total reactions (Children)	EEF (Children)	Element only (Children)	Total instances (Total children exposed)
Active						
Andy eats carrot	10 (4)	8 (4)	18 (5)	11 (5)	-	29 (5)
Andy says hello when touched	-	-	-	1 (1)	1 (1)	2 (1)
Andy says raining on me	-	4 (3)	4 (3)	4 (3)	15 (9)	23 (9)
Carrot changes cloud colour	6 (4)	5 (3)	11 (5)	14 (6)	4 (3)	29 (9)
Child discovers apples are stackable	2 (1)	-	2 (1)	-	-	2 (1)
Child discovers berries can't change cloud colour	-	-	-	1 (1)	-	1 (1)
Child discovers carrot properties	-	2 (2)	2 (2)	-	-	2 (2)
Child discovers white tube properties	3 (2)	-	3 (2)	-	-	3 (2)
Child picks carrot	1 (1)	2(1)	3 (2)	17 (7)		20 (8)
Child reacts to seeing carrots (all picked)	1 (1)		1 (1)	-	-	1 (1)
Child sorts two apples stuck together	1 (1)	1 (1)	2 (1)	-	-	2 (1)
Child tries apple in wrong colour tube	1 (1)	1 (1)	2 (2)	1 (1)	-	3 (3)
Flowers "dance" when pressed	1 (1)	-	1 (1)	1 (1)	-	2 (2)
Snail eats carrot or strawberry	5 (2)	4 (3)	9 (4)	4 (4)	-	13 (5)
Passive						
Andy says can you pick carrot	-	-	-	6 (6)	3 (3)	9 (8)
Andy takes apple sorting turn	2 (2)	7 (5)	9 (5)	13 (7)	-	22 (8)
Andy takes cloud turn	4 (4)	4 (4)	8 (5)	12 (5)	-	20 (10)
Child reacts to changed apple colour	-	1 (1)	1 (1)	-	-	1 (1)
White tube changes apple colour	1 (1)	1 (1)	2 (2)	8 (5)	2 (2)	12 (7)
Grand Total	**38**	**40**	**78**	**93**	**25**	**196**

Table 8.12: Child reactions to inherently novel elements discoverable in addition to ordinary game play

System error	Initiation	Non-social reaction	Error, no reaction	Total
Active				
Andy steals apple from child	2			2
Andy won't eat item	1		1	2
Carrot cannot be picked	1		1	2
Carrot stuck to cloud, hovers in sky			1	1
Child accidentally activates menu mid-game, changes activity		1		1
Difficulty with touch-screen	66	45	48	159
Object bounces rapidly between floor and ceiling	8	9	7	24
Objects stuck together	11	4	1	16
Touch-screen fails completely			2	2
Subtotal	89	59	61	209
Passive				
Andy apple mistake: throw 2 apples at once	1		1	2
Andy arm malfunction during cloud turn	2	3	12	17
Andy does not say intro/close seq.	1	2	6	9
Andy makes growing activity mistake: plant already finished	2		2	4
Andy makes growing activity mistake: wrong side	1	1	2	4
Andy steals apple from child		2		2
Andy stops in middle of turn		2		2
Andy will not take turn	2	3	14	19
Apple appears before Andy has arms in position	2			2
Apple ricochets into tube		1		1
Apple stuck to Andy's hand		1	1	2
Apples stuck mid-air	2	1	4	7
Cloud changes colour without object touch		1		1
Error message box appears	1		1	2
Fireworks do not ocurr		1	3	4
Fireworks reward appears in middle of activity			1	1
Fireworks reward delayed		1	1	2
Object disappears, or lost offscreen	1	1	3	5
One apple becomes 4 in basket			1	1
Sound effect fails to play: Various		4	5	9
White box appears over Andy	2	2	2	6
Subtotal	17	26	59	102
Total	**106**	**85**	**120**	**311**

Table 8.13: Error-related DR pairs and error-only instances

Perceived discrepant aspect	Primary initiations	Non-social reactions	Total UD
Active discrepancies			
Andy can eat a strawberry	1		1
Andy eating is "not working"	3	1	4
Andy eats again after "not working"	1	1	2
Andy will not accept a given object		1	1
Andy's utterance appears rude or inappropriate		1	1
Apple will not go in other colour's tube	4	2	6
Apples cannot be re-attached to tree	1		1
Butterfly does not react to touch		1	1
Cannot remove apples from basket	1		1
Carrot is "upside down"	3	2	5
Child cannot feed Andy a flower	2		2
Child successfully takes turn after many failed actions	3	1	4
Cloud stops working after all items are grown	3	4	7
Multi-touch cannot be used	2	1	3
Plants do not grow when touched, child surprised		1	1
Putting carrot on Andy head (novel action)	1		1
Snail eating is "not working"	1	1	2
Something about an object is different or "broken"	2	2	4
Something wrong with activity transition		1	1
Surprised to get carrot after 2 strawberries	1		1
Touchscreen stylus is "not working"	1		1
Aactive subtotal	30	20	50
Passive discrepancies			
Andy doing nothing, child seems to think he should	2		2
Andy takes a turn after long inaction		1	1
Apples "missing" from the basket	1		1
Nothing happening, something should be	12		12
Object perceived missing, perceived to have disappeared	2	1	3
Researcher made mis-statement or mistake	5		5
Something different about Andy cloud turn	1		1
Strawberry turns back into apple when sorted	2	1	3
Three flowers are the same colour (child surprised)	1		1
White tube produces different apple colour than expected	1		1
Passive subtotal	27	3	30
Total	**57**	**23**	**80**

Table 8.14: Children's reactions to perceived discrepancies

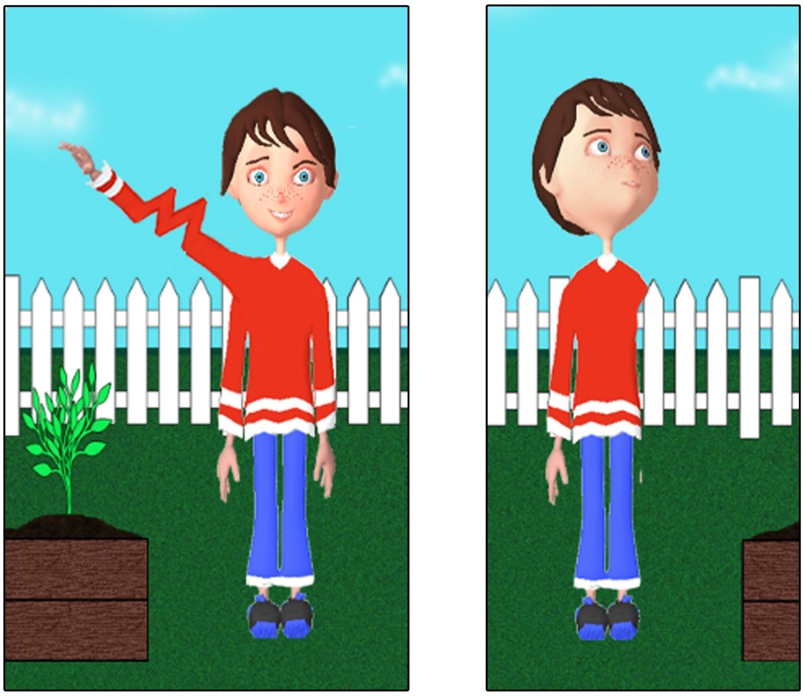


Figure 8.5: Two system errors regarding Andy's arm movement, when taking turns in the growing games. *Left*, Andy is stuck with three arms (two images displaying at once). *Right* No arm image displays, and Andy is stuck looking up.

several children's reactions to the *same* DDO are presenting in tandem, as a comparison. Each example or group of examples is accompanied by brief commentary.

Kristina reacts to a novel purple flower Kristina[11] finishes growing a flower for the very first time and experiences some novel effects around the flowers. A purple tulip blooms and a fanfare sound plays. As soon as the flower blooms, she stops touching the cloud.

> Kristina: *Looks at the flower and claps her hands three times.*
> Kristina: *Immediately turns to look at the researcher, simultaneously using her forefinger to point at the flower. She rapidly shifts her gaze from the researcher to the flower, and back, and comments* "Purple". *She looks back to the screen.*
> Researcher: *Says encouragingly* "It's a pretty flower!"
> Andy: *Smiles* "Good job!"
> Kristina: *Looking at flower, repeats quietly* "purple".

[11]Consent was not given for this child's image to be used in publications or presentations.

Per pre-study information from her teacher, clapping is Kristina's usual way of showing happiness and excitement. She likes the flowers and the fanfare, and wants to share them. Here, both the researcher and Andy give timely and positive encouragement in response to her sharing. This is an excellent example of a child non-socially reacting (via clapping) and then using a combination of behaviours to initiate about an inherently novel game element.

Omri reacts to an apple turning into a strawberry In session 3, Omri[12] put an apple in the white tube and it became a strawberry. Despite many suggestions from the researcher, he refused to try the white tube until session 3, offering various bits of explanation along the lines that it was the wrong colour for the apples. This was the *only* time he tried the tube, and he does it spontaneously. He seems quite interested in the result, but not necessarily very happy about it. The child's comments may not seem socially directed from their description, but Omri speaks in a noticeably different manner when directing his comments to someone else, versus narrating to himself. His self-directed speech is quiet and extremely indistinct, often almost incomprehensible, but he is louder and enunciates much more clearly when he is trying to communicate with others. This is a case where the child's typical pattern of behaviour is very important in determining whether specific actions should be considered to constitute initiations.

> Omri: *Pokes the strawberry with the touch-screen stylus and says loudly and distinctly* "Is turn it into a strawberry!". *He lifts it up then drops it to let it bounce.*
> Researcher: *Agrees enthusiastically* "That's right!" *pause* "It's magic!". *As she speaks, Omri begins sorting a yellow apple.*
> Omri: *Waits for the yellow apple to appear in the basket, then begins moving the other apples and berry on the ground, trying to pick one up. He says something quiet and incomprehensible except for the words* "have to".
> Omri: *After another pause, he picks up the strawberry and now says very clearly* "we have to make it back into apple". *He sorts it into the red tube. The berry goes in. When it re-appears in the basket at the bottom, it shows up as a red apple.*
> Researcher: *As apple appears, she asks* "what happened?"
> Omri: *He is looking at the basket. After a pause, he says* "Is gone back to apple."
> Omri: *Comments* "Now I happy."

Omri's DR pairs were annotated as *surprise*, not because the tube has a different-than-usual effect (he has no information about any other effects), but because the apple

[12]Consent was not given for this child's image to be used in publications or presentations.

unexpectedly becomes an object with no prior place in the apple game. Omri does not seem to like this result, and takes it upon himself to resolve it. It is surprising that he overtly *predicts* that the red tube will change the berry back to an apple. Other children sorted the berry, but most did not notice that it changed back in the basket. It is also unusual for Omri to volunteer that this new situation makes him happy. This final comment is an example of a *secondary initiation* that is part of an ongoing interaction, but introduces new information or a change of topic. Omri adds a new—rare—piece of information about his emotions.

Omri's experience with the strawberry is a prime example of an instance where a child did not necessarily *like* a discrepancy he detected, but the situation still motivated several initiations. He remains calm and engaged throughout this episode, trying to "fix" the rogue strawberry and expressing happiness when he succeeds. Children noticing or interacting with DDOs is a reasonable proxy for interests and attention, but should not necessarily be considered a proxy for what they *enjoy*.

Bradley initiates following Andy's end-of-activity disappearance At the end of the flower game in session 3, Andy tells the child "Well done, you grew all the flowers!". About two seconds later, he disappears with a pop, leaving the center of the screen empty.

> Bradley: *He is reaching for the screen when Andy disappears. He stops, looking at the centre of the screen.*
> Bradley: *After a pause, he turns to look at the researcher, apparently seeking information.*
> Researcher: *Suggests moving on to the next game (A missed opportunity to scaffold an interaction about the disappearance).*

Bradley showed far less interaction with the researcher than any other participant, so his turn to the researcher is particularly noticeable. His reaction is shown in Figure 8.6, in three steps. This DDO was equally surprising and worthy of a social reaction the second time it was seen. Playing through the games again, Bradley saw the same DDO approximately 8 minutes later.

> Bradley: *He is leaning forward with his elbows on the table, appearing to lose interest in the game. Andy disappears with a pop.*
> Bradley: *He sits up sharply, taking his elbows off the table and turning immediately to look at the researcher.*
> Researcher: *Meets child gaze, then looks to screen and makes an exaggerated expression of surprise, possibly mirroring the child's own expression*

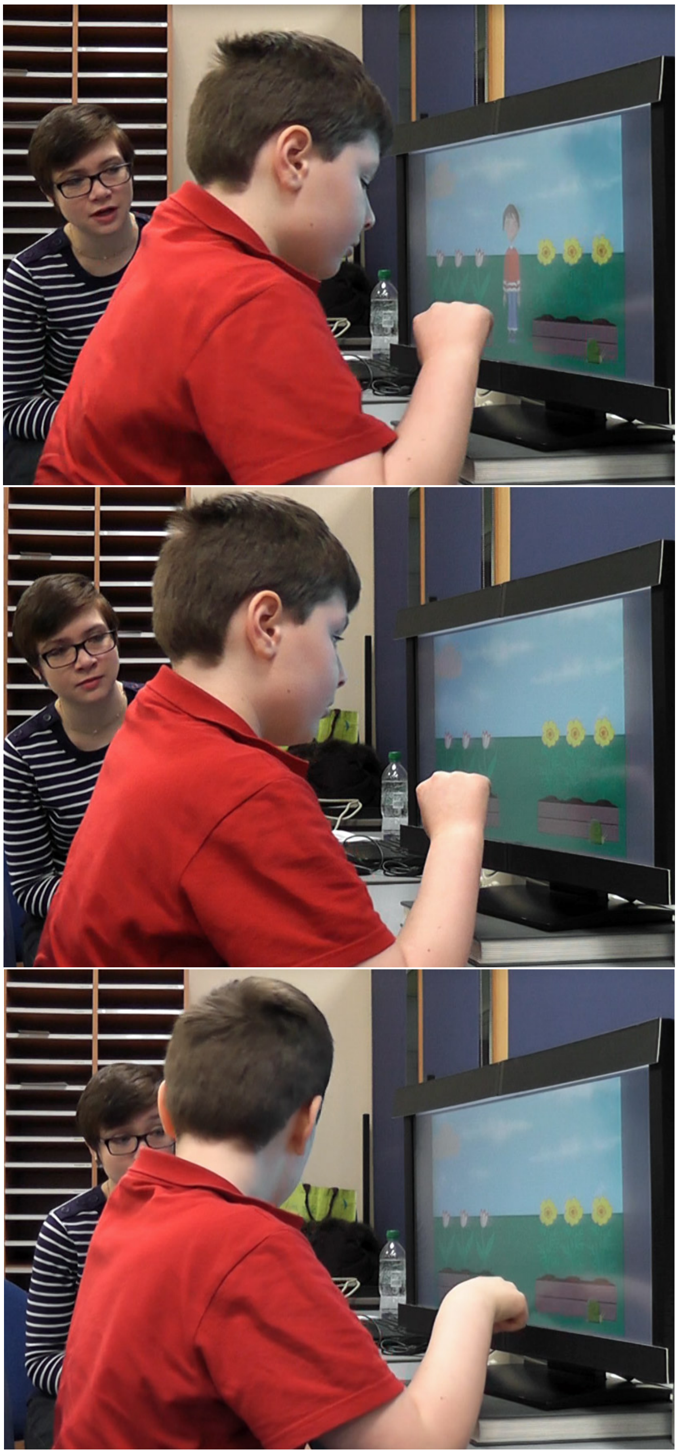


Figure 8.6: Bradley reacts to Andy's disappearance for the first time, looking at the screen in apparent surprise and turning to seek information from the researcher (social referencing).

at the time, which is not visible in the video.
Bradley: *Reaches for bottom of screen, his indication that he wants to access the menu and move on.*

Bradley seems, if anything, even more surprised when Andy disappears again. He literally sits up and takes notice, and turns to the researcher immediately (less than 2 seconds after the disappearance). Perhaps he requires less processing time upon seeing this event a second time. Unfortunately, the researcher does not do a good job of taking advantage of these relatively uncommon initiations, and trying to scaffold a further interaction. Admittedly, Bradley's responsiveness to language was patchy throughout the study, and it was difficult to know how to encourage additional interaction.

Three children react to the missing cloud At the start of two different altered mini-games, the cloud is not present on-screen at the activity start. It is absent long enough for the child to notice it (about 8 seconds), and then glides on-screen from the left. Andy delays his opening sequence until after the cloud has appeared. Let us consider how three different children reacted to this DDO:

Activity opens
Alfred: *Child is looking at the snail, then looks up to top left of screen, reaching for the cloud. He takes his hand away, looking at the empty space* (Figure 8.7.
Alfred: *Leans his face very close to the empty space, grabbing the top of the screen and leaning on it. He says* "Where's the cloud?" *The cloud enters almost immediately afterwards*
Researcher: *Comments* "There it is, it came back."
Alfred: *He leans slightly away from the screen and moves his hand down from the top to use the cloud. He does not make any other reaction.*

Activity opens
Eshan: *Immediately reaches to where cloud should be and starts pressing down. Nothing is there.*
Eshan: *He asks loudly* "Wh— Where's the cloud?" *pulling his hand away from the screen and continuing to look at the empty space.*
The cloud enters after a short pause
Researcher: *She says enthusiastically* "Oh! There it is. It came back." *She also points but this does not seem to be in the child's line of vision.*
Eshan: *He reaches for the cloud and tries to use it before it is even all the way on-screen. He does not comment or show any other reaction.*

Activity opens
Arthur: *He looks to where the cloud should be. He briefly touches the*

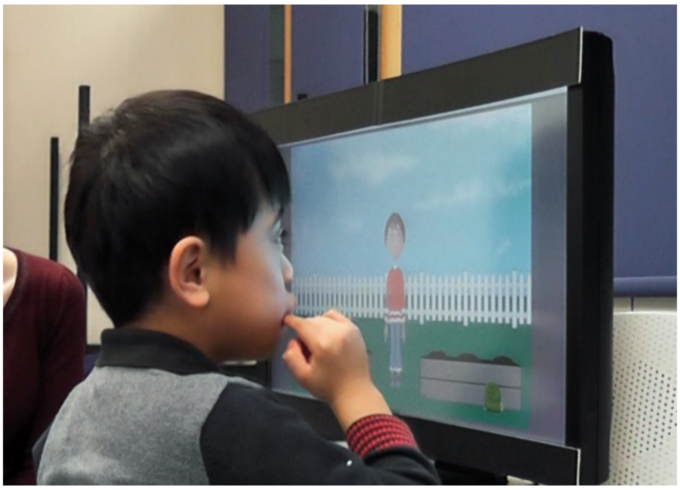


Figure 8.7: Alfred looks for the cloud. A few seconds later, he leans on the screen, with his face close to the area where the cloud chould be.

> *empty space, then spends several seconds looking around the screen, moving his hand above it. He swipes insistently several times at the point on-screen where the cloud should be. When this has no result, he sits back, propping his elbow in the table and leaning on his hand* (Figure 8.8). *He looks at the screen and waits for something to happen.*
> *The cloud enters*
> Arthur: *Still leaning, he reaches for the cloud with the other hand as it moves on-screen. Then he sits up and switches hands to use the cloud.*

Alfred and Eshan have quite similar reactions, obviously noticing the cloud is missing and seeking information from the researcher. While neither of them look at her, neither child tends to do so when he converses. Eshan (like Omri) tends to speak much more loudly and clearly when he is directing his speech to others than when he is self-narrating. Arthur, who infrequently uses verbal language, reacts to the same event non-socially. While he chooses to wait and see if something happens, rather than asking for information, he shows the same initial pattern of reaching for the cloud and visually searching the screen.

This DDO (missing cloud) was quite effective at eliciting reactions across the participant group. The cloud's timed re-entry was meant to *resolve* the DDO, meaning that the child would not be blocked from pursuing the activity goal. This strategy of using a timer to create a *temporary* missing object appears to have been quite successful, and the timing was adequate to allow notice, but not lead children to become frustrated because they could not do the activity.

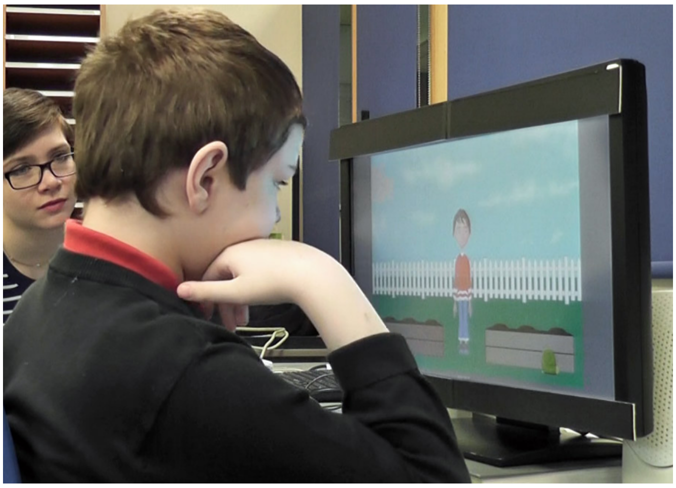


Figure 8.8: Arthur leans on his arm, waiting for something to happen. The cloud is just beginning to come on screen, top left (not easily seen in this image).

8.5 Discussion

8.5.1 Child communication

All 10 children with autism in the current analysis reacted to a range of discrepant aspects, across the games sessions. These children represent a range of autism severity, linguistic ability, and cognitive ability. Collectively, they made 650 reactions to 526 unique discrepant aspects, a mean of 65 reactions to 55.5 different aspects. A full 62% of these child reactions were initiations to a social partner, with the child who had the smallest proportion of initiations (Bradley) still showing nearly 40% social reactions. Across the group, the balance of social and non-social reactions is closely comparable to that in the original ECHOES data (study 1; 64% of reactions to a social partner) suggesting that the current games have met an important goal with respect to re-creating the pattern of social interactions seen in ECHOES.

Across all children, approximately one in four primary initiations about discrepancy continued on to one or more subsequent initiations (*secondary initiations*) about the same aspect. Some children continued to secondary initiations far more often than others, from 4.17% for Arthur, up to 42.11% for Omri. What this means is that for children such as Omri and Martin, a sizeable proportion of their initiations to a partner were the first step in a longer interaction, mixing child initiations, partner actions, and child responses to the partner (response data collected but not analysed in this study). Other children, such as Arthur, Bradley, and Connor, tended to have “one-off” initiations, perhaps a single child initiation and response from the partner, or a child initiation followed by a few short partner-initiated exchanges.

The researcher was the target of 90% of child initiations, significantly more than to the virtual character. This was a shift from the original ECHOES, where 30% of the initiations were directed to Andy. A likely cause of this change is that ECHOES-Andy was far more sophisticated and dynamic than the Andy in the current game. He was rendered in 3D, walked about the screen, could bend, turn, and jump. An AI planner determined his activities—though this was slow, and responsible for many of Andy's apparent "mistakes". ECHOES-Andy also greeted the child by name (pre-recorded) at the start of each session. Overall, the character in *Andy's Garden* seemed like much less of a social partner because he *did* less. It is encouraging that children so clearly recognised this, and directed themselves to the partner who was most likely to be responsive and helpful.

8.5.1.1 Child affect

Andy's Garden appears to have created a positive play experience for the participating children, with only 10% of child reactions appearing to show mildly negative affect (i.e. frustration, discouragement, annoyance, disappointment). Other reactions were positive or neutral, with approximately twice as many neutral as positive reactions. Often, children may have been in a more positive or negative state, but did not show it (or not in a way understandable to a person with limited knowledge of that child). Some children in the study appeared largely impassive, which is not unusual for children with autism. There were no child reactions that could be considered highly negative, or that caused an attending school staff member to intervene or express concern. No participants became extremely agitated, or exhibited challenging behaviour such as pushing or yelling. Even when children became loudly frustrated, they were evidently motivated to continue their game play, rather than abandoning an activity. Some children requested help many times in a row before finishing an activity. Almost all of these negative reactions involved the touch screen interface, which proved a point of difficulty for almost every child—again, similar to what was seen in the original ECHOES study. The touch screen is further discussed in 8.5.3.3.

8.5.1.2 Rates of reaction and child characteristics

Children played with *Andy's Garden* for varying lengths of time, most often determined by how long the child wished to continue. Children had a mean of 47.5 minutes of video data (total= 475 minutes), though this varied from 23 to 73 minutes[13]. Considering children's rates of reaction or, specifically, rates of initiation, gives the best estimate of "how much" children reacted to discrepancy, as it accounts for their varying time with the games. Across the group, children reacted to something discrepant a mean 1.4 times per minute. This does not mean that they necessarily detected a new discrepant aspect every 1.4 minutes: this rate includes secondary initiations (i.e. multiple initiations about the same aspect). Considering initiations only, the group initiated to the partner about discrepancy a mean 0.89 times per minute. Bradley was by far the least active initiator, with a rate of about once every six minutes. This seems consistent with high SCQ score (34) and his observed behaviour elsewhere in the school context, i.e. appearing frequently withdrawn or "in his own world". However, the significant result here is that this child did seem quite engaged by the game play, and *did* regularly react and initiate about it, even when he did so much less frequently than his schoolmates. It meant that when Bradley did initiate, it was particularly noticeable and striking. One such example was described in 8.4.3.

A valuable finding here is that *all children*, even those at the higher end of the SCQ range, were initiating throughout their game play. There does not seem to be a SCQ "cut off" above which the environment was unsuccessful at engaging children and motivating initiations. This is good news for future designers designing for groups of children with mixed ability.

8.5.2 Design

While there were many possible DDOs and errors in Andy's Garden and individual children's lists of DR pairs look quite different, there are points of overlap between all (or the majority) of children. Overall, the design appears to have successfully engaged all children at some point. As predicted by the high-level design principles (Chapter 6), it is effective and apparently non-disruptive to include frequent, widely varied DDOs in a game design when designing for a group of children whose particular profiles and interests are not known in advance. Further analysis of existing data from this study and Study 1 might generate additional design profiles (or design personae), that could

[13] Campbell, who is missing game session 3, is the child with the shortest session.

be used as a tool to narrow down future design choices, towards those most likely to engage and motivate particular sub-groups of children.

8.5.2.1 Active and passive discrepancies

The interactivity of discrepancies seems to be an important factor for most children, with active discrepancies appearing more likely to elicit reactions, rather than just notice. This may be for several reasons: developmentally, it may be difficult for many of the young participants to self-inhibit, wait, and watch DDOs: Andy's sequence of actions to take a turn (correct or incorrect) is a good example. They may even struggle to pause and look around the screen, rather than reaching in immediately to the first thing they see. Whether children have good self-inhibition or not, DDOs that children can only encounter through direct action may simply be more interesting. The experience of discovering or revealing something may be attractive and motivational in itself. This could be further explored in future work.

8.5.2.2 Discrepancy and activity goals

The pattern of child reactions across discrepancy sources suggest that many children found DDOs and errors particularly salient when these prevented or delayed them from carrying out their goals. This is most notable in the case of the difficult touch screen interface (see 8.5.3.3) that motivated so many initiations to the researcher, requesting help to pick up and move objects. There are other examples, however, such as the designed DDO in which the cloud is initially missing at the start of the growing activity, and errors where objects were stuck together, or Andy took apples from the child. Conversely, DDOs and errors that were unrelated or weakly related to activity goals seemed less likely to be encountered at all (where these were discoverable) or to motivate child reactions.

Some children, such as Eshan and Omri, seemed quite focused on the activity goals and adhered closely to those goals as they understood them. In addition to reacting, sometimes frustratedly, to delays, they were often extremely *resistant* to exploring the environment and trying things that did not seem clearly related to the task, such as putting apples in the white tube. This limited the number of DDOs to which they were exposed. As a counterexample, Alfred was highly exploratory, and perceived discrepancies when some of the actions he expected to work (e.g. feed Andy a flower, Andy accept apple from child) were not possible in the game.

In future designs, DDOs that have a direct relationship to activity goals and object functions may have the best chance of being noticed by children, and motivating their reactions. While some children will be interested in non-functional and non-goal-related changes and additions, some subset of children will not be. A design should not focus too heavily on features of this second type. At the same time, designing DDOs *only* related to goals might be unwise and increase the chance of frustrating or alienating a goal-focused child, who could feel like the whole game is “not working” and he should stop playing.

8.5.2.3 Inherent novelty, and getting used to the games

In the ECHOES system, most children were highly reactive to exploring the Magic Garden and encountering the objects, sounds, and rewards in different activities. Based on this pattern it was expected that initially seeing and exploring *Andy’s Garden* would motivate many DR pairs. Some novel elements (i.e. fireworks reward, Andy eating a carrot, carrot changing cloud colour) motivated a sizeable number of reactions (e.g. 11-18) across at least half the participants. Far more elements motivated only 1-3 reactions each, over 1-3 children. Overall, children playing with *Andy’s Garden* seemed less inclined to exclaim, laugh, and share as they initially discovered the garden[14]. With the exception of the fireworks reward, those novel elements motivating the most reactions were all discoverable elements, in addition to ordinary play. As discussed above, discovering something may be inherently more interesting to children than passive exposure.

Children’s developmental age may potentially play a role in their preference for novelty, or at least their tendency to become excited about it or share with others. The group playing *Andy’s Garden* was, on average, developmentally older than the ECHOES group by 17 months: a sizeable difference, in early childhood development. Developmentally older children may be less motivated by new things simply because they *are* new, or potentially may be inhibiting their reactions.

8.5.2.4 Sound and DDOs

Sound effects appeared to “under perform” compared to expectations in the design phase. Analysis of discrepant aspects in the original ECHOES (Chapter 6) suggested that adding and changing sound effects might be interesting to many children, and

[14]It is possible that their familiarity with the researcher may have affected this: in ECHOES, children had much more time to get used to the researcher, prior to the start of the technology sessions.

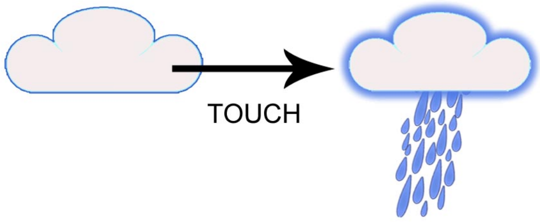


Figure 8.9: Feedback when the child touches the cloud

that it would be worth increasing their presence in the current design. Overall, this prediction was not supported. Many of the routine sounds, such as flower fanfares, did not seem to excite most children when first presented, or later altered. There were many sound effects in *Andy's Garden*, which often attracted no notice at all. A few of the exceptions are illuminating, with respect to the role of sound in the game environment.

Consider the designed DDOs "cloud sound missing" and "tube sound missing": the first resulted in zero reactions or EEF across all children (from at least 25 instances), the second resulted in 9 reactions and 5 further instances of EEF from eight children. The difference may lie in each sound's importance as feedback. When a child holds her finger on the cloud to take a turn, it lights up with a blue outline, produces rain, *and* makes a rain sound (see Figure 8.9). When sorting an apple into a tube, the sound is the *only* feedback for several seconds, until the apple appears in the bottom basket with an additional sound effect. Many children appeared to want this confirmation of a correct action, and stopped their action and watched for the apple to appear in the basket, instead of moving on to the next apple immediately. In the absence of a sound, some looked around for the apple they just sorted, as though it had actually dropped or bounced away. For the cloud, sound is not needed as feedback, and it is unclear if any noticed it was even missing.

Finally, one sound-related DDO was unexpectedly very successful: children made a number of initiations about Andy's mis-statements. Here, Andy called items the wrong colour or mis-named them, such as "It's a red flower!" when the child grew a yellow flower, or "Look, a yellow apple!" when the white tube produced a strawberry. It had been predicted in the design phase that most children would not notice these because they were relatively subtle, language-based DDOs. Here, three children (Alfred, Kristina, and Leanne) initiated 10 times about these, with an additional child making non-social reactions. Sometimes children were quite indignant about these statements, leading to multiple exchanges with the researcher about Andy's mistake. It is worth noting that two of these children, Alfred and Leanne, were identified by teachers as particularly talkative, and also liking *others* to talk.

8.5.2.5 Go big or go home

The current game design generally chose to be conservative in designing DDOs, with respect to how much they affected the games. For example, the DDO "Andy disappears at activity end" was debated as potentially too upsetting, or confusing when Andy reappeared in the next activity. Actually, it was quite successful, motivating 10 initiations from six different children, and EEF from three more[15]. Children were surprised or entertained, but not remotely upset. Conversely, some of the DDOs included to test the lower bound of noticeable DDOs were clearly *too* subtle: No children even noticed the removal of the garden fence or change of one planter's colour in session 3. These were both non-interactive objects, and unrelated to the activity goals. Other borderline DDOs, noticed or reacted to in only a few instances, included changing the growing behaviour of plants and substitution of non-feedback sound effects (cloud sound, flower fanfares, fireworks reward).

Overall, some of the most successful DDOs (in terms of EEF and reactions) were those that had the greatest degree of incongruity. Events like "some apples do not fall from tree", "snail added to garden", or changed flower colours may all violate established rules and patterns of the game, but still make perfect sense in terms of the child's knowledge of the world. An apple or carrot turning into a strawberry is *much* more incongruous, and violates the world's rules as well as the garden's rules. Children reacted accordingly. The white tube changing apples' colour or producing two apples seem like they should follow the same pattern, but do not. At least for producing two apples, it is possible most children were developmentally not yet able to register this as an "impossible" event. Perhaps the researcher referring to the white tube as "magic" also altered children's expectations about it.

Considering incongruity does not completely explain the pattern of children's reactions, however. If big apples make sense, so should big carrots—but five children reacted 15 times to giant carrots, and only one child reacted once to giant apples. As discussed elsewhere, the child's direct action and the sudden "revealing" of the giant carrot may both have played a role in making this DDO more motivating than the giant apples, which are present when the activity starts. Or, perhaps the apples just weren't

[15] Only 9 children were exposed to this DDO, as Campbell missed session 3.

big enough!

It is still unclear how disruptive DDOs can be before children find then genuinely unmanageable. However, it is worth keeping in mind that most children with autism (like other children) consume a steady diet of cartoons, books, and games stuffed with outlandish and impossible events. Indeed, literature on children's media could prove a valuable source of ideas. Future designs in this area could probably add bigger, brighter, more disruptive DDOs with no ill effects for most children. The main area of caution would be to consider how DDOs impact children's progress toward activity goals. This seems to be the area in which there is the lowest tolerance for change.

8.5.3 Methodological issues

8.5.3.1 Ethics

This study aimed to practice "most-informed consent" with child participants, giving each child as much information about the proposed research participation as appears feasible for the him or her, and having a "consent conversation" with the researcher. This conversation was meant to communicate to children that participation was voluntary—they could leave when they wished without being in trouble or hurting our feelings. They could say yes or no to us taking a video, and to us showing the video to "other people who make games for children" (see Child Consent Form, in Appendix H).

Child information sheets and consent forms were sent home with parent information, and several came back with child signatures. One child, Leanne, told researchers about looking at the forms with her parents. However, attempting "consent conversations" with other children at the school was not notably successful, for several reasons:

- We attempted to do this when children arrived for their first session, in the room where the study would take place and where the computer was present.

- Prior to the first sessions, teachers and staff had already told children they would be working with us today, and sometimes also told children they would be using the computer and/or playing games.

- Prior to the first sessions, some teachers put pictures of the researchers or games on children's visual schedules, in the same manner as they do with required school activities.

By the time we tried to approach children about consent, they had already received a strong message that this was a "regular" school activity, and thus implicitly not-optional, and not up to the child when it would be over. We then tried to speak to them in what turned out to be a distracting context: in the presence of a computer with which they expected to play. Even when it was hidden with a cloth, children knew the computer was in the room and were extremely eager to get to it. Most had also been told that their next activity was the computer, or "playing with Andy"—not stopping to have a talk with the researcher! It was not a good time to get children's attention for a complex topic. As a result, there was not a consent conversation for most participants in the way that had been envisioned. All agreed that they wanted to play the games and none expressed concern about (or even interest in) the video camera, as a few children with autism had spontaneously done in previous studies (Alcorn, 2010; Frauenberger et al., 2013). However, it is unclear to what degree children understood their participation as a choice.

The author is still of the firm opinion that consent conversations are a good ethical practice, and can work with children in this age and ability group, with appropriate planning. In the future, it would be critical to speak to children *before* teachers have announced the research session as one of the child's daily activities, and to meet with them on a day prior to the first session—when they are not expecting to imminently play with the computer. It would also be essential to speak to teachers about this consent philosophy, and work with them to identify a way to present the activities to the child such that s/he knows they are a choice. Teachers, staff, and researchers could then work together to deliver a consistent message.

8.5.3.2 Familiarisation

The time invested in pre-study familiarisation activities and teacher discussions appeared to pay off in the experimental sessions. Aside from Campbell, whom teachers warned would have difficulty with transitioning to any new activity, the participating children displayed little or no anxiety about leaving class to participate in the experimental sessions with the researchers. Indeed, many of them seemed to approach the sessions eagerly, as an exciting thing to do. The teacher discussions also proved very useful for data analysis, as teachers had explained common behaviour signals for many children (e.g. Kristina claps to show she is excited). These kinds of behaviours, which teachers can "read" easily, are difficult for researcher to pick up in the short space of class observations but can be very important for interpreting a child's communicative intent or affect.

8.5.3.3 Issues with the touchscreen interface

As noted in the procedure, the touch screen used in this study was a smaller, less sensitive substitute screen. Children may have experienced it as *especially* insensitive because of their daily experience with smart boards in the classroom. Several accompanying staff members volunteered that these are highly sensitive to touch. From researchers' in-class observations, the smart board programs also mainly use single touches over dragging.

All participants struggled with the touch screen to varying degrees: some had difficulty grasping how to do the actions but then had little trouble afterwards. Others struggled persistently throughout the sessions. The apple game was a particular problem, as it required apples to be "lifted" and dragged. Maintaining consistent touch and dragging were both difficult for children, but unlike the cloud, the apples would fall if touch was released. Partway through the study, the physics of the apple game were modified such that apples would fall more slowly, allowing children a chance to recover and catch them. This does not appear to have substantially lessened children's difficulty.

Motor control is also a factor in touch screen use, and a known difficulty for many young children. One child excluded from the study, Colin, had incomplete data largely because of the extent of his motor difficulties. As several other children in the study used paper cut-outs in the familiarisation drawing activity due to their difficulty with controlling pens, it is unsurprising that they also struggled with the touch screen. What is, perhaps, surprising is the degree to which all children *persevered* with the games, in spite of the interface.

Children's persistent touch screen difficulty—and the multiple apparent sources of this difficulty—calls into question the common view in the children's technology community that touch screens are the easiest option for young children. The experience of this study and Study 1 suggests that this is not universally the case. Touch screen interfaces are considered in more detail in Chapter 9.

8.6 Summary of study 2

This study sought proof-of-concept that a game could be designed to deliberately feature situations and events that young children with autism would perceive as discrepant with their expectations, and that these discrepancies could motivate children to initiate to social partners. Ten children with autism (M=8, F=2; mean developmental age 5 years) completed three sessions[16] with an experimental touch-screen game, *Andy's Garden*, that included three mini-games or activities (flower-growing, sorting apples by colour, and carrot-growing). Children were videoed playing for an average of 47.5 minutes each, often requesting to repeat the activities again after completing them the first time. Annotation of children's emotions and their willingness—even excitement—to attend sessions and complete activities suggest that they found their participation not only manageable, but positive and enjoyable.

Children reacted—and initiated—to discrepant aspects at a high rate, and in a predominantly positive way. Individuals and the group showed a fairly even balance of reactions about the discrepancy subtypes (novelty, surprising non-events, and surprising events), and detected a range of discrepancies related to system errors and child perception, in addition to the designed DDOs and inherently novel game elements. As predicted in Chapter 7 (game design), individual children had some overlap in the aspects to which they reacted, but outside of a few particularly popular DDOs, their interests and styles of play varied widely. Children seemed particularly attracted to hands-on discovery of the environment and of DDOs, and were engaged rather than upset by the designed DDOs that changed the environment the most, or were most incongruous. Between DDOs and system errors (notably, a difficult touch screen interface with low sensitivity) the participating children tolerated quite a high level of disruption and change during the game play. Universally, they were engaged enough to try to persevere past problems and errors, often seeking help from the researcher in order to do so.

The results of this study allow us to answer its overall questions affirmatively: it is possible to motivate children's communication by including deliberately-designed DDOs in a set of games. The basic strategy for "creating" discrepancy, proposed in Chapter 6, appears to have been very successful. A session with a "normal" game version was used to establish children's expectations and familiarise them with game mechanics and goals, and then two altered game versions attempted to violate these

[16] One child completed only two sessions.

expectations in various ways. As in the original ECHOES video data, the data from *Andy's Garden* suggests that detecting discrepancies between their expectations and the environment, including encountering novel game elements, appears to be "worth communicating about" for young children with autism. It is an extremely encouraging result that *all* children in the group followed the same basic pattern of detecting and reacting about discrepancies, although their rate of reaction varied considerably. The group results were not "driven" only by a few highly responsive children. Even the *least* responsive participant, Bradley, made 11 spontaneous initiations in his hour of game play, only one of which was about trouble with the touch screen. He is a child with a high SCQ score (i.e. larger number of autistic traits), who often appeared to be withdrawn from his surroundings, and whose BPVS score suggests intellectual disability in addition to his autism. In short, he is a child who—on paper—would not be considered a particularly promising study participant. Engaging a child with this profile and motivating him to initiate seems more significant than it does for some of the children who teachers already identified as particularly talkative or interactive, in pre-study discussions.

Overall, Study 2 appears to provide a clear proof-of-concept that including DDOs in a design is a promising strategy for encouraging children with autism to spontaneously initiate, a developmentally critical skill. Future work in this area can begin to fine-tune design strategies, and branch out to examine a wider range of technologies, participants, and discrepancy-detection opportunities.

Chapter 9

Discussion and conclusions

9.1 Introduction

The previous chapters have pursued an investigation of a previously unobserved phenomenon, children with autism communicating about *discrepancy*, and sought to describe, categorise, and transfer it to new contexts. This final chapter reviews the research questions successively addressed in this work, summarising the answers to each question and highlighting how these fed into the next stage of work. The discussion then takes a step back, to consider issues and limitations of the current body of work. It identifies specific contributions of the work, and broader connections to the rest of the autism and technology field. Several specific pieces of future work are suggested, including additional theoretically motivated analysis of child communication in the existing datasets, and new work that would extend and revise the current concepts of sameness and discrepancy-detection opportunities (DDOs). Finally, the chapter and the thesis conclude with a short summary of key findings from this work.

9.2 Research questions revisited

The introduction and research approach chapters summarised the line of investigation pursued in this thesis as a sequence of five research questions, with each question being successively identified based on the outcomes of the previous questions. Together, they represent the process of closely understanding and describing a novel phenomenon in one context (ECHOES), and then going through a process of transferring ECHOES insights and lessons to a new context *(Andy's Garden)*, aiming to re-create a similar pattern of positive child communications. The rest of this section steps through the questions individually, drawing out their relevant findings that have been presented over the course of the thesis. The discussion of RQ5, which asks "does it *[Andy's Garden]* work?" also includes more general commentary on the extent to which the study 2 results provide proof-of-concept for DDOs as a design strategy.

[RQ1] What is happening when children with ASC were observed to communicate about novel and apparently surprising aspects of ECHOES? What is the nature of this phenomenon (eventually labelled as discrepancy)?

This is a qualitative question, answered by describing and identifying initial examples of the phenomenon that (later in the thesis) became known as a *discrepancy-reaction pair* (DR pair; see Chapter 1). These follow a pattern of children noticing something in ECHOES that they found new or different in some way, and reacting to it in a positive and frequently social way. Even from the early examples, the emerging concept of discrepancy was always concerned with multiple components: something in the world, and a child reaction.

The main work pertaining to this question is reported in Chapter 4, as the earliest stage in the iterative process of developing the discrepancy-reaction taxonomies and annotation scheme (see Table 4.1). Phase 0 (Initial observations, data immersion) looked across the ECHOES corpus to identify more examples that seemed to follow the pattern of the initially observed examples. A key product of RQ1 was a clearer formulation of the topic under investigation, and proposal of tentative codes by which to capture the child reactions and the environmental aspects. These provisional, working concepts are a pre-requisite for RQ2, which asks about the extent of the discrepancy phenomenon in the ECHOES dataset.

[RQ2] Is discrepancy represented by a few rare but striking instances in the ECHOES dataset, or are there sufficiently widespread examples to merit further investigation and form a basis for categorisation?

This question was addressed in Phase 1 of the taxonomy and annotation development (Exploration; Initial, non-systematic coding), in Chapter 4. It took the provisional concepts from RQ1 and used them to guide the review and open coding of ECHOES participant videos. The main goal of this surveying process, as RQ2 indicates, was to try to go beyond the vivid, initially observed examples, and to see whether similar instances did in fact appear throughout those children's data, across different learning activities and days of play.

Answering this question was not about exact quantification—no target number or frequency of instances was set, that would indicate "enough" data for further analysis.

What was important was to determine if the initial examples were the *only* examples, and how any further examples were distributed. Phase 1 of the taxonomy and annotation development clearly answered this question, identifying *many* more examples of the phenomenon across children and ECHOES sessions. These followed the general pattern of original examples, but represented a greater diversity of reactions and system events. The answer to RQ2 was yes, that DR pair examples were extensive enough to proceed to labelling and categorising instances, and developing an annotation scheme. If this surveying had *not* identified examples beyond the initial observations, or very few more, then this would have necessitated a very different approach to subsequent analyses.

[RQ3] Given that instances of discrepancy appear repeatedly across ECHOES sessions and participants, how can we describe and categorise them? Which features or sources of information need to be included in order to capture the nature of these instances?

a In what ways do children react to discrepant aspects?

b What things or events in the environment appear to be a source or cause of discrepancy (i.e. discrepant aspects)?

Having cleared the practical hurdle of RQ2, the remaining phases of Chapter 4[1] document the iterative development of taxonomic categories, and an annotation scheme that applies these categories to data. As has already been documented at length, the answers to RQ3 underwent substantial revision across the rounds of analysis[2] before reaching their final form.

An important part of the answer to RQ3 is the *point of reference* for description and categorisation: the individual child's experience and knowledge (Section 4.3.2). Discrepancy is not an objective and independent "thing" out in the world, but exists and must be labelled relative to a particular child. Our description of discrepancy must thus also include the child. The unit of description and analysis is thus the *discrepancy-child reaction pair* (DR pair; see Section 4.3.3). When we describe discrepancy, we need information about the situation in the world, and about the child's reaction. In early versions of the taxonomies and annotation, these pieces of information were enough. However, the final discrepancy concept (a *core category*) refers to something else: a

[1]Phase 2 (Proof-of-concept analysis; used middle-order coding), Phase 3 (First-draft version of taxonomy, prior to core category) and then Phase 4 (Final taxonomy with core category, final annotation scheme). See Table 4.1.

[2]Some of these shifts are summarised in Table 4.2.

particular *relation* between the child's knowledge and the environment, in which the child detects a subjective inconsistency between the two. When we categorise discrepancy into subtypes, we are referring to different types of inconsistency between child knowledge and the environment. *Novelty* means that a child currently has no expectations about some part of the world; it is unpredictable on the basis of his current experience. *Surprise* means that child's expectations about a known part of the world have been violated in some way—the expectations are subjectively inconsistent with the world. In order to make this relation judgement, we must draw on other sources of information about a child's past experiences and about the environment. This makes sense, given the situational nature of discrepancy. This additional information about environmental events and specific child behaviours (i.e. those that would determine whether something qualifies as an initiation) are recorded, though not explicitly labelled and categorised, during the annotation process. The discrepancy taxonomy, or, more specifically its implementation as annotation scheme categories, provides the main answer to RQ3, and how to describe and categorise discrepancy. The child reaction taxonomy sets out the answer to sub-question RQ3-a, about how children react to discrepant aspects. Children may react either by initiating to a social partner, or reacting non-socially. They may react to the same discrepant instance multiple times. Currently the specific behaviours within child reactions and the function of their initiations are not described in detail; this is reserved to future work.

RQ3-b, about environmental causes of discrepancy, is addressed outwith the discrepancy and reaction taxonomies (Chapter 4) and their application to ECHOES data (Chapter 5). The exact "causes" of child reactions, or aspects about which they reacted, are not included in the taxonomies, although that environmental information *is* collected as a part of the annotation scheme (see Figure 4.4). This information was analysed to provide input to the design process (Chapter 6, Section 6.2.1), yielding the answer that almost *anything* can be a discrepant aspect, due to the element of child perspective involved, and the unplanned nature of the non-novel discrepancies in ECHOES. There were 112 *types* of aspects that children detected as discrepant, grouped into 19 mutually exclusive *aspect groups* (see Appendix E). The takeaway message from this is that there does not seem to be a small number of aspects (or even a few related groups of aspects) that appear mainly responsible for the discrepancy phenomenon. Many different aspects motivated DR pairs, and children seem highly individual in those that they found most noticeable or interesting. This is positive partly because it supports a view that the discrepancy phenomenon is not specific to the exact contents and behaviours of ECHOES, and that a future design in this area would have many options open to it.

[RQ4] Is it possible to create discrepancy and discrepancy-contingent communications on purpose? How might this be achieved?

a Per the concept of discrepancy as a *relation* or *state of affairs* between the child and the environment, how would a designer try to facilitate this phenomenon?

b What features of the ECHOES context appear to have contributed to a *motivating but manageable* experience of discrepancy?

RQ4 asks about the possibility of creating discrepancy and related communications, on purpose rather than by accident (as in ECHOES). This question focuses more on theory and planning—how would we do it? RQ5 asks whether the strategy is successful in practice. Unpacking the concept of discrepancy gives the answer RQ4-a. If discrepancy is subjective and particular to individual children, then we cannot talk about creating discrepancies in the abstract. Chapter 6 explains that if discrepancies are *situations* that motivate certain child reactions, the designer is trying to create more of these situations. S/he will insert objective change, novelty, or inconsistency, with the goal of the child recognising and reacting to some of these as *subjectively* inconsistent. The designer creates opportunities for discrepancy detection (DDOs), rather than discrepancy itself. ECHOES naturally provided opportunities of this type, either through novel elements that were a deliberate part of the design, or through system errors.

Chapter 6 developed the working hypothesis of a *motivating but manageable* environment as a way to account for children's positive reactions to discrepancy, in the context of the need for sameness. It appears that encountering discrepancy can be motivating and interesting to children, but *also* emotionally manageable, when these experiences are balanced out by a larger quantity of sameness. This concept of motivating manageability gives a very high-level vision for a future design. RQ4-b used further reflection, lessons learned, literature, and Study 1 data in order to hypothesise about what "went right" in ECHOES, and contributed to this productive balance. These hypotheses were presented as a set of six high-level design principles intended to support transfer of "motivating but manageable" DDOs to new designs: establishing and maintaining *integrity* of the environment and activities, ensuring *flexibility* and *resolvability* of child-system interactions, and offering wide *variety* of DDOs at an appropriate *frequency*. DDOs should pose ambiguous *opportunities* for communication, rather than demanding specific behaviours or communicative forms. The guidelines were framed by a set of *assumptions* about the goals and intended use of a future design that includes DDOs. These are also a tool to support transfer, because they identify further elements that (per current understanding of motivating but manageable interaction) should be *similar* between the sending (ECHOES) context and the receiving (new design) context, in order to reproduce the same types of interactions.

[RQ5] Guided by design principles rooted in empirical work with ECHOES and autism theory, can designed *discrepancy-detection opportunities* (DDOs) successfully create a *motivating but manageable* experience with new participants, in a new set of games?

- **a Will children be motivated to spontaneously initiate social communication about the designed DDOs?**
- **b Will children's game-play experience be emotionally manageable, per their affect, utterances, and behaviour during games sessions?**
- **c How do specific design decisions appear to affect children's communication and their overall game-play experience?**

RQ4 asked whether transfer of discrepancy-contingent communications appeared possible, and what factors might have contributed to the original interaction pattern in ECHOES. This research question is about putting the transfer into practice: having designed a custom receiving context in the form of new games, does this design create the effects we have tried to transfer? Chapter 7 reported the implementation of the design principles in a suite of new games, *Andy's Garden*, which sought to balance designed DDOs with sameness in order to create a *motivating but emotionally manageable* experience. Chapter 8 reported a proof-of-concept-scale evaluation study with the new games, and new participants. The qualitative description and quantification of child behaviours from this study directly provide, or enable reflection on, answers to the RQ5 sub-questions.

The main goal was always for the *overall effect* of the design to re-create a pattern of positive, discrepancy-contingent interactions, including the effect of system errors and child perceptions. Overall, the design of *Andy's Garden* appears to have successfully transferred key insights from ECHOES, and does show a remarkably similar pattern of child-system and child-adult interactions. Of 650 child reactions to discrepancy[3], 409 (62.92%) were directed to a social partner, predominantly the researcher. If asking "does the design work?" and "does it create the intended overall effect?" the answer is yes to both.

RQ5-a asks whether the designed DDOs[4] were able to motivate child initiation, as they were intended to do. The answer again is yes; 74 out of the 294 discrepancy-related primary initiations were motivated by designed DDOs (25.27%). A further 74 initiations were motivated by the inherently novel elements and actions that make up the basic design and game play of *Andy's Garden*. All ten participants initiated about designed DDOs, though their level of apparent interest varied. The 59 additional non-social reactions to designed DDOs also illustrate the design's success, because they represent DDOs attracting child interest and notice, even if they did not motivate communication. Given the short total duration of each child's game play—and the fact that designed DDOs were present only in two out of three sessions— these results are an encouraging proof-of-principle for the basic concept of DDOs proposed in answer to RQ4-a. It *does* work for researchers to design situations that they think children will find inconsistent with their previous experiences in an environment. At least part of the time, children were "taking up" these opportunities by spontaneously reacting to them.

One additional, important support for the success of DDOs as a concept is the increased "share" of child DR pairs in *Andy's Garden* that were examples of surprise–slightly over a third of all reactions, and slightly less than a third of unique discrepant aspects. This is a substantial gain from what was seen in the original ECHOES data. ECHOES had many inherently novel elements and a range of system errors (the source of many non-events), but there were relatively few "naturally occurring" episodes perceived as surprising *events*, rather than non-events. Examples such as the character making an activity mistake were memorable for children and the researcher, but relatively rare. Trying to deliberately create surprising events was a priority for the new design, as a test of designed DDOs as a strategy.

The results reported in Chapter 8 suggest that children found their overall experience with *Andy's Garden* to be emotionally manageable, in addition to motivating (RQ5-b). The simplest piece of evidence is that there were no episodes of serious emotional dysregulation at any point during the study. All participants were willing,

[3]Including primary initiations, secondary initiations, and non-social reactions.

[4]As opposed to DDOs that might not be designed, e.g. due to system errors, or child perception.

even eager, to return on multiple days and "play with Andy again". Looking at child affect data in more detail, the vast majority of child reactions appeared to have a positive (29.69%) or a neutral (59.23%) valence, as judged by children's verbalisations, facial expressions, gestures, and movement (see Section 8.4.1.3 and 8.5.1.1). Instances of identifiable negative affect were all mild and relatively uncommon in the data set, comprising 66 out of 650 reactions, or 10.15% of the total. Almost all of these negative reactions involved the touch screen interface, which proved a point of difficulty for almost every child–again, similar to what was seen in the original ECHOES study. The various measures to balance discrepancy with sufficient sameness appear to have been successful. What is not yet clear is the role that different *types* of sameness (e.g. in-game routines, physical environments, repeated dialogue) may play in an overall manageable experience. This is discussed under future work.

Difficulties with the touch screen and subsequent help requests are one specific example of the effect of specific design decisions on children's interactions (RQ5-c). Choosing to use the touch screen and to combine in-game physics with object lifting and dragging (in the apple game) directly caused many of the difficulties that children experienced, and frequently resulted in their initiating to request help.

RQ5-c is a question that has produced some specific insights, but a greater number of likely future investigations. It is the type of question that can only produce a provisional answer. It is difficult to summarise the effects of other specific design decisions—such as DDOs—because there were so many individual decisions, and frequently little overlap between the aspects to which children reacted (as also seen in the original ECHOES). However, some specific decisions have been analysed in the Chapter 8 discussion (Section 8.5.2)[5]. Further work with new designs will be better placed to examine the effects of specific design decisions.

9.3 Issues and limitations

9.3.1 Small scale of studies

This thesis freely acknowledges that both of its studies are on a proof-of-concept scale, using two deliberately similar technologies and focusing on children in a similar range of ability. Understanding fewer participants and contexts, in more depth, has been very

[5]For example, use of character "mis-statements", and extent to which child reactions were affected by DDOs being congruent (or not) with children's knowledge of real-world objects and gardens.

valuable in these early-stage investigations of discrepancy. However, it does limit the conclusions that can be drawn. As yet, it is only hypothesised that children's interest in and positive reactions to discrepancy would appear across a broader group of children with autism, or in dissimilar technologies. Or, in other words, that discrepancy is a general motivator, in the same way that need for sameness is a general characteristic of autism. More cases would need to be added to the current body of evidence—expanding participants, technologies, and contexts of use—in order to treat embedded DDOs as a general and portable design strategy for motivating initiation.

Other ASC-tech literature gives us reason to believe that further investigation *would* provide this evidence, and support the concept of discrepancy-detection as an inherently interesting experience that can motivate communication. Several other authors have included very intriguing anecdotes in their publications of incidents that are meant to primarily illustrate other points, but follow the same type of discrepancy-reaction pattern as seen in the two studies here. Consider this description from van Rijn and Stappers (2008), in commentary on their participants' sensory preferences in the LINKX project, which used "smart objects" to aid association of words and objects in the child's home:

> For Robbert, LINKX' sounds were very important. When he linked a block to a speech-o-gram on the cat's litter box, the block's lights turned red, but the sound was not produced. After a moment, Robbert instructed the block to "say litter box." For Jakob, LINKX' colours were very important. Each time a block changed its colour, Jakob enthusiastically pronounced the colour "red," "green, or 'blue." He did not mind the two times that the sound failed. (p. 6).

Robbert's behaviour seems closely equivalent to the ECHOES participants who prompted or scolded Andy for his mistakes. His *intention* in this case seems social, even if his 'partner' is not. Jakob, interested in another aspect of his experience, is not disturbed by the objective violation of the speech-o-gram's usual behaviour—just as many ECHOES participants peacefully ignored objective violations and errors that did not interest them.

Another example comes from a social robotics project, and demonstrates not only apparent discrepancy-detection, but a wonderful example of a technology mediating communication (discussed more in Section 9.5.3). Robins and Dautenhahn (2006) report a case in which Jack (age 7), plays a game where a small robot imitates his body movements. Unknown to him, it is actually controlled by the experimenter:

> ...it just happened, on one occasion, that the experimenter unintentionally

> moved the opposite arm of the robot. Jack giggled and mentioned (to the robot) that this was wrong....The experimenter then introduced more deliberate mistakes, and Jack's laughter and affection directed at the robot grew. Then an important point arrived when Jack realized that the experimenter was operating the robot from his laptop and that it was him who was making the mistakes, so it then became a game between the experimenter and Jack.... after each mistake that the robot made in mirroring Jack's movements (which were deliberately introduced by the experimenter), Jack turned to the experimenter laughing saying 'mistake', 'mistake'... Jack was using the robot as a mediator to indirectly interact and play a game with the experimenter. (p. 648)[6].

Both of these examples are from technologies extremely different than ECHOES: a tangible interface using blocks around children's houses, and a social robotics project. Both of them also would absolutely be annotated as discrepancy-reaction pairs under the current scheme. There are doubtless more anecdotes of this type hidden in ASC-tech papers, and even more that ended up on the cutting-room floor as interesting, but not part of a paper's main story. These examples lend support to the hypotheses pursued in the current work, that discrepancy may be generally interesting and motivating to children with autism, *and* that it can be a manageable and positive experience when other circumstances are right.

9.3.2 Second-coding of study 2 data

Study 2 reports DR pair presence, discrepancy subtypes, and child reactions, identified according to the same scheme and procedure used in Study 1. Chapter 5 reports second coding on a subset of Study 1 data, producing strong inter-rater agreement on the presence and labelling of DR pairs. However, it would be best practice to also second-code a subset of study 2 data. This could be particularly important if seeking to publish stand-alone study 2 results in a psychology-oriented venue, where second-coding is an expected practice (far more than it is in HCI, for example).

It would also be best practice to conduct second-coding and inter-rater agreement calculations for the design-related annotation labels introduced for study 2, namely the categories of children showing *evidence of expectation-formation* (EEF) about elements of the environment, but falling short of behaviour that qualified as a non-social reaction. This is the new label that requires the most judgement on the part of the annotator. Having confidence in judgements of non-social reactions versus EEF would be-

[6] This is an abbreviated version of a much longer passage, which is worth reading in its entirety, and also includes pictures of the robot and imitation game.

come more important in a more controlled, comparative study that sought to measure the efficacy of *particular* DDOs in producing reactions. Some of the other design-related annotation labels probably do not need second coding (or have it as a lower priority) because they are much more objective. If a coder knows how a technology *should* work and what it contains, it is fairly easy to identify erroneous system behaviour. With knowledge and documentation of the design, it is also quite objective to label the source of DR pairs: a designed DDO, a system error, inherent game elements or child perception (any DR pair that is not one of the first three!).

9.3.3 Touch-screen interface

Study 1 chronicled the many problems that ECHOES participants had with the touch-screen interface, particularly the intersection of touch actions with in-game physics. *Andy's Garden* tried to mitigate some of these issues, for example by *substantially* slowing all object physics (so that a falling item could be caught and the action recovered) and making objects slide over each other instead of bouncing off. These changes appear to have made only minimal differences: the exact same problems appeared with all Study 2 participants, even the oldest children. Repeated touch screen difficulties was a source of frustration in both studies. However, even where children showed negative affect, they also demonstrated high motivation to continue playing with both *Andy's Garden* and ECHOES, often repeating actions or asking for help many times because they were determined to finish. The fact that children showed frustration, especially with the touch screen, should not lead us to conclude that their overall play was unmanageable. As noted, identifiably negative reactions were by far in the minority, and many of them involved children constructively seeking help from the researcher. Sometimes they sustained multiple-turn interactions as they needed to explain what help they wanted. Given this overall pattern of interaction, the touch-screen actually *supported* the design goals of *Andy's Garden*. The touch screen represents a level of desirable difficulty in interacting with the environment. Where technologies aim to foster communication—or cooperation, or collaboration—a touch screen may create a useful challenge. It can motivate repeated communication, if an activity is sufficiently motivating, and if children cannot reliably use it alone. If a child can always successfully complete every game action herself, there is much *less* to communicate about.

More "traditional" interfaces might not offer a better alternative for young children. In a study comparing child performance on point-and-click versus dragging tasks with a mouse, young children (5-6 years) struggled more, were slower, and made more errors with dragging actions (Joiner et al., 1998). Inkpen (2001) had similar findings. The real problem for young children might not be the input method, but the pressing-and-dragging action itself, coordinating motion and constant pressure on a screen or button. Anthony and colleagues summarise that, when dragging, "children make less stable movements, have difficulty maintaining contact with the screen, and make more input errors overall than do adults" (2013, p.158). On balance, the examples of ECHOES and *Andy's Garden* in conjunction with literature on mouse use does not suggest any winning interface for young children. However, minimising the need to drag objects, especially in conjunction with physics, might substantially improve ease-of-use in future designs.

9.4 Reflections on methodology and epistemology: Pragmatism, "subtle realism", and constructivism

Chapter 3 discussed the overall research approach and methodological choices in the work, identifying the work as broadly qualitative, contextualist, and driven by pragmatism. The goodness-of-fit or "dovetailing" (Bryman, 1988) between the information produced by a method and the research questions to be answered was treated as paramount. Pragmatism approaches individual methods as tools, rather than as representing commitments to specific philosophical positions. Reviewing the current, completed, work indicates that the methods chosen consistently "dovetailed" well with the research questions, and have been successful tools in producing new knowledge that helps to solve the 'problem' posed by the research questions. The commitment to considering behaviour in context has been particularly important for understanding the current phenomena of interest: individual child reactions to discrepancy must be understood in the context of that child's "history" of game interaction, and in relation to actions taken by the social partner and the technology. In this sense, the current position has been an effective one, by which the author continues to stand. However, further and more specific commitments can also be claimed: to some extent they are implicit in the work already conducted, and can now be put forth more clearly and strongly prior to framing a programme of future work.

Reflection on the completed work reveals a consistent epistemological position underlying the specific methods chosen, and the overall logic of the research. It most resembles the broadly positivist approach of Hammersley's "subtle realism" Hammersley (1992a,b)[7][8]. This position acknowledges an external, objective world—including a social world—that exists independently of individuals and their subjective experiences. However, it combines this with the more interpretativist caveat that we may only *know* this reality from a certain perspective. The observer (or rather, the researcher) has values, theories, pre-existing knowledge, and other features that affect *what* is observed and interpreted. In some cases, the researcher is a further degree removed, interpreting the participants' or respondents' interpretations of the world. When we give accounts or descriptions of reality, these are *representations*, not reproductions, and there may be multiple valid descriptions of the same phenomenon Hammersley (1992a)[9]. Under this view, objectivity (of accounts of reality) and neutrality of the researcher remain important ideals toward which to work, even if they cannot be reached (Snape and Spencer, 2003).

Some kind of realism was a necessary commitment within the work, in order to engage in practices such as the second coding and calculation of inter-rater agreement present in Chapter 5 of this thesis. The overall line of of investigation, already revisited in this chapter, also has strong realist implications and engages in hypothetico-deductive activity. The set of research questions effectively ask about whether certain observations [of an external world] by a particular observer are part of a pattern or phenomenon that "really exists". To what degree does the initial *account* of a phenomenon (children reacting to discrepancy) correspond to the external reality? Is this account *plausible* to my research community (or communities), and is it *relevant* in the sense of producing a certain type of outcome (Hammersley, 1992a)?. At the same time, the implied position is consistently aware of the more interpretivist part of Hammersley's subtle realism. As discussed throughout the thesis, the concept of discrepancy itself is fundamentally about individuals' subjective *interpretations* of reality, though in many cases it is most concerned with the limited, sandbox-style reality of a computer game. Discussions of the discrepancy annotation scheme and process are clear that they are about *interpreting* the reality of child behaviour and experience, and representing it through labels and categories.

[7] As also usefully discussed in Seale (1999), Silverman (2011) and Angen (2000).

[8] Several authors who discuss subtle realism and place it in a wider context of approaches to research, or specifically to qualitative work, overtly identify it with pragmatism (Seale, 1999; Snape and Spencer, 2003).

[9] Though he is quick to specify "to this extent, I agree with the relativists, but I must stress that I do not accept that there can be multiple, *contradictory*, yet valid accounts of the same phenomenon" (1992a, p.199).

Subtle realism appears to fit together quite naturally with a more constructivist view of research as a practice or set of knowledge-production processes. Here, the researcher—with values, biases, theories, and experiences—actively selects some things as data (but not others), and actively constructs accounts of reality, mediated by his or her own perception of it, and in some case also the perception of participants or respondents. The ultimate accounts or constructions of reality are conjectural rather than certain. Like subtle realism, this constructivist view of research is implied throughout the current thesis, but was not stated overtly at the outset. For example, in Chapter 4 there is discussion of *constructing* discrepancy as a core taxonomic category, and video annotation as a process of *selecting* some segments as analysable. To varying degrees, Chapters 4, 5, and 8 all discuss the researcher/author's experience with the study settings and child participants, and how these may be relevant to the analyses undertaken and the accounts constructed.

The position of subtle realism, a constructivist view of research activity, and pragmatic methodological choices has served the present research programme well. Future work that is similar in its questions and participant group could continue on this footing. However, a revised position may be required if later work chooses to address research questions by asking participants to give their *own* accounts of their understanding and experiences. For example, an evaluation might feasibly involve interviewing older and more verbal children with autism about their understanding or experience of a game design with DDOs. Interviewing or other techniques that directly elicit participant experiences deal with the "social world" far more directly than does the current work, and they must address and reflect on different underlying assumptions. For example, to what extent might the participant and researcher co-construct concepts like "surprise" during an interview? Would it even make sense to assume a social world independent of individuals, especially when some of the individuals have autism, and researchers and/or readers may not? To what extent should the researcher's construction of an account, and choice of terminology, try to reflect participant experiences? These questions might all need to be addressed, and the basic methodological position revised, in future work. Some of these questions may not have easy answers in existing literature, or parallels in current studies, due to the conjunction of participants' young age and their autism.

9.5 Broader implications and contributions

9.5.1 Research contributions

9.5.1.1 The discrepancy concept

The central contribution of this thesis is the discrepancy concept itself, or more precisely, children with ASC reacting to discrepancy. This phenomenon began as a collection of unusual, theoretically interesting observations, and has been systematically explored and defined using an iterative, inductive process. The concepts thus produced were repeatedly tested against the dataset through constant comparison, and by "implementing" them as video annotation labels and procedures. The result is a concept, discrepancy, that can be operationalised in terms of child-environment interactions in such a way as to identify new examples, in new participant data. Discrepancy has been further tested by second coding, using the annotation scheme. This demonstrated that a person with comparatively limited knowledge of the original video context could identify and label DR pairs at a rate of high agreement with the more knowledgeable researcher-annotator. DR pairs represented a small proportion of the overall data. Moreover, the discrepancy concept has also been *applied*, as part of a new design. The results of study 2, in which children *did* communicate socially and positively about designed DDOs, are a realisation of the potential implied by the original observations of discrepancy in ECHOES.

The methods pursued in this thesis represent a contribution in themselves. More specifically, the contribution is the *collection* of methods that were employed at different points, to pursue the concept of discrepancy from initial observations, all the way through to deliberately applying it in a new technology design and assessing the effects. It represents a pragmatic, systematic, and principled approach to methodological choices, that may be a useful example for other researchers. This is particularly true where they are developing ASC-tech projects that do not represent a version of face-to-face interventions or support—that is to say, are not building directly on specific prior work. As the corpus of ASC-tech research grows, we will likely see more phenomena like discrepancy: novel, potentially useful interactions that have only a distant or tangential precedent in existing literature. The current work offers a model for "building up" from such observations, systematically investigating, testing, and applying their insights.

The line of investigation to develop and test the concept of discrepancy has produced a number of valuable contributions along the way, representing tools for other researchers to use in understanding and applying this concept. First, there are the discrepancy-reaction taxonomies, and their implementation in the annotation scheme. With an eye to future re-use or modification by other researchers, the annotation scheme described in-text is one part of a detailed "package" of information. The complete annotation manual is included as Appendix C, with more specific guidance and worked examples. This was used for training the second coder. The second coding itself is reported in detail, including coder training, test-coding and "calibrating" discussions between the first and second coders, and appropriate choice of statistical tests. Taken together, there is a complete package of information to encourage others to use and build on the current annotation scheme.

9.5.1.2 Design contributions

With respect to applying discrepancy in design, there are several contributions that again represent a methodological template, here for explicitly supporting transfer between ASC-tech contexts. Together, the overall working hypothesis of *motivating but manageable* interaction, the six design principles[10], and the detailed implementation information in Chapter 7 all function as a complete package of information about embedded DDOs as a new technology design strategy for ASC. They give a clear account of how different sources of information contributed to the principles, and document the additional decision-making required to realise the principles in a specific design that uses DDOs to motivate communication.

The literature on technology design and autism is growing rapidly, but still small compared to the literature about typically developing children and adults. There are few sources that make overt design recommendations for autism-focused technologies, or aim to provide general guidance to other researchers and designers interested in this group. The most widely-known is **?**, sharing recommendations based on the LINKX tangibles project, described in 9.3.1. Many projects and papers in the area of design and autism are most concerned with *participatory design*; they are papers in which the key outcome is a novel or newly-adapted methodology for meaningfully and safely involving children with autism in design processes (e.g. Frauenberger et al., 2013; Benton and Johnson, 2013; Malinverni et al., 2014, 2016; Millen et al., 2012) and interpreting their inputs (Frauenberger et al., 2012b). Others give accounts of participatory design

[10] An initial version was published in Alcorn et al. (2014).

processes as part of a project (e.g. Keay-Bright, 2007, 2008; Porayska-Pomsta et al., 2012; Grawemeyer et al., 2012). Many of these papers provide a wealth of "lessons learned" about successful and unsuccessful design for children with autism, but do not have transferability as an overt goal.

The current work broadly agrees with the recommendations in **?**, who are concerned with slightly younger children. They also emphasise the importance of creating structured situations (mostly discussed in terms of *integrity* in the current work), but also allowing children control over interactions (related to the design principle of *flexibility*). The authors also highly recommend the use of sensory rewards such as music, pressure, and vibration, and they acknowledge the need to accommodate different sensory preferences and child special interests (related to the current principle of *variety*). The success of sensory rewards and the variability in child preferences was also borne out by the current game designs. **?** also recommend that designers capitalise on and further develop the strong visual memory often evident in this group. In many respects, the concept of establishing and (then violating) child expectations is very reliant on these kinds of memory skills, and confirm them as a strength.

The six design principles presented in Chapter 6 add to the small corpus of design recommendations currently available in this area. It is positive that they broadly agree with **?**, given that they were developed in a completely different context (screen-based media with exploratory play, versus tangibles for word learning). They also go beyond **?** by giving a much clearer account of their development, with additional connections to literature and theory. While the principles are focused on designing a particular type of interaction—social interactions around discrepancy detection—their broad agreement with **?** supports the idea that they can also give wider insight into successful design for ASC.

The higher-level goal of *motivating but manageable* interactions may be particularly useful as a tool for other designers. It has an implicit heuristic that designers must consider the balance or interplay between these two factors, and also must overtly try to create manageability for this group (whether this is in the form of sameness and routine, or managing the multiple demands on users). The goal is domain-independent, and applicable regardless of what the specific technology tries to do, or the exact user characteristics. For a new design brief or situation, the goal may actually be used as a set of starting questions about what might *motivate* children to engage at some point in an interaction, and what they are presently being asked to *manage* (e.g. task demands, sensory input, language processing...). The value of this heuristic is that it *is* general, and focused on the experience of the child as a technology user. It also is concerned with the *overall effect* of a design as experienced by the user, and how individual elements (including the context of technology use) combine to create this experience.

9.5.1.3 Discrepancy as an educational strategy

Thus far, discrepancy and DDOs have been discussed in terms of motivating social communication, and designing technologies that seek to help children with ASC practice these key skills. However, DDO-type situations could have wider application in the field of educational technology. They could be a useful tool for learners both with and without autism, at a range of ages and in a range of domains. Nothing about the concept of designing DDOs is specific to autism, games, children, or exploratory interaction. The basic idea is to introduce objectively new, different, or erroneous events or pieces of content into a known context, with the goal of an individual finding them *subjectively* inconsistent, and taking some kind of action. Identifying a subjective inconsistency between current knowledge or expectations (what we might call a child's "mental model") and the state of the world requires a child to make a comparison between the two[11]. As initially discussed in Alcorn et al. (2013a), the process of children identifying discrepancies in games is arguably equivalent to learners identifying and correcting"bugs" in their own academic knowledge by comparing themselves to an expert source[12]. Both require an individual to consider current information, events, or procedures in light of what he already knows (i.e. in relation to his mental model), and to conclude that something is unknown or inconsistent in some way.

Discrepancy-detection in the current context shares an important underlying mechanism with the strategy of *erroneous worked examples* used in a number of existing educational technologies, namely intelligent[13] maths systems for older TD learners (e.g. Tsovaltzi et al., 2012; McLaren et al., 2012; Adams and McLaren, 2013). These have found that including erroneous worked examples, among other types of problem solving, can successfully increase application of metacognitive skills, and also positively affect learning. The task of identifying an erroneous step in a worked math example and the opportunity to detect Andy making a sorting mistake both rely on the same type of *metacognitive* process (Flavell, 1979). There must be that comparison of

[11]There is no claim that children make a conscious or deliberate comparison, in the discrepancy-detection examples discussed in this thesis.

[12]Such as a piece of software, or even an expert-like character within the software.

[13]I.e. Artificial intelligence-driven systems such as those in Adaptive Learning Environments and Intelliegent Tutoring Systems literature; see Woolf (2009) for a general introduction.

a "mental model" to the world, whether the current domain is mathematical facts and rules, or the contents and rules of *Andy's Garden*. In both cases, a child must reflect upon her own knowledge and *representation* of the world (Hacker, 1998). This is the "thinking about thinking" at the heart of metacognition.

The existing educational technologies that use erroneous examples (e.g. Tsovaltzi et al., 2012; McLaren et al., 2012; Adams and McLaren, 2013) differ from the current examples of discrepancy-detection in several important ways:

- The learners' age and support needs;
- The maths systems teach specific domain content and focus on explicit problem-solving, rather than being exploratory;
- The maths domain content is propositional and academic;[14]
- System-side maths errors are objective, and may be either procedural or factual[15];
- System-side maths errors are "announced" to the learner: finding and correcting them is set as a specific exercise and teaching opportunity[16];
- The goal of including system-side errors, in the form of erroneous worked examples, is to strengthen learners' domain knowledge and metacognitive skills.

In several respects, those systems and the current examples of discrepancy-detection represent the end points of several different continua of learner characteristics, types of content, and types of interactions. In each case, instances in which a system, a perceived "expert", is objectively incorrect can galvanise learners to metacognitively reflect on their mental models and take some kind of action. What the work on discrepancy-detection adds to the current pedagogic picture is that objective, "system-side errors" can be effective at getting learners to engage in metacognition and take relevant action *even when they are not set as an explicit task*. The "errors" do not have to be prepositional; they can have to do with relationships, patterns, and sensory content. Where the content is appropriate, the strategy is also relevant for young learners who may have limited language and no reading ability.

[14]The domain content in ECHOES and Andy's Garden, such as it is, is the content and behaviour of the game environment.

[15]In ECHOES and Andy's Garden, "errors" tend to involve constitute disrupted cause-and-effect relationships, or alterations of sensory or temporal aspects of the environment.

[16]Versus the unannounced and ambiguous opportunities present in ECHOES and Andy's Garden, where no corrective action is *required*, although it may be possible.

This research picture, with erroneous worked examples for older learners at one side and the current work at the other, suggests several possible avenues of future application. First, if discrepancy-detection is indeed generally interesting to children with ASC, as well as requiring metacognition, it may constitute a strategy for engaging children with ASC with technologies that teach academic skills. These are also a key area for ASC-tech (e.g. Pennington, 2010) and a major concern of parents and practitioners. Employing discrepancy in academic-focused software may mean erroneous-example type problem-solving tasks, further "unannounced" DDO-like situations, or something in-between, like alerting children that a virtual tutor or character "sometimes makes mistakes" and the child "might want to check and see if he is right".

For typically developing learners, unannounced DDOs may be a useful complementary teaching strategy to erroneous worked examples, where error correction is set as an explicit task. More work would be required to determine how to deploy DDOs adaptively in an intelligent tutoring system, and also how to address instances in which learners "miss" opportunities, or find objective correct content to be subjectively incorrect (like many of the *perceived discrepancies* documented in Chapter 8). While these situations make little difference with exploratory, non-propositional content and communication goals, they pose more of a problem with factual content and academic learning goals. DDO-like situations might best be used to challenge and engage learners who already have good domain knowledge and are equipped to identify and address system-side errors.

The original concern, of DDOs as a way to motivate communication, may also have broader applicability in educational technologies that are meant to be used collaboratively, either by students directly working together or as part of a community of learners (e.g. an online course). Encountering unexpected, novel or erroneous content may *also* motivate other groups of learners to turn to another person (or course message board, etc.) and communicate about system content. Help-seeking communications, when they appear, represent another important and difficult set of metacognitive behaviours (Gall, 1981) that continue to challenge many learners even through adulthood (Aleven et al., 2003).

9.5.1.4 Contributions related to autism and ASC-tech

As discussed further in the following section, technologies can be tools for learning about the characteristics and needs of autism. The analyses in this thesis suggest a new and useful view on the *need for sameness* that is central to the diagnosis of autism.

This view is more complicated, and less absolute than that that sometimes emerges from the literature. While many practitioner-authored or practitioner-targeted sources may already treat the need for sameness more flexibly than does the research literature (e.g. Wall, 2010; Howlin, 1998), these often focus on specific situations and examples, rather than *general* ways to think about this issue. The idea that we can counterbalance difference with substantial sameness is a very useful frame for thinking about a range of technological and non-technological situations. Sameness is a slider bar rather than an on-off switch. Considering multiple *sources* of sameness (e.g. within an activity versus in the child's surroundings) is also useful, because it suggests means by which we may provide—or subtract—sameness, to impact an overall experience. The work presented here, while on a small scale, suggests the potential of using technologies to learn more about what the need for sameness means, and how it can operate in the world. This is discussed more under Section 9.5.4.

Throughout the thesis, discrepancy has been repeatedly connected with the initiation of new communications. Initiation is a difficult skill set to target, especially due to the requirement that children demonstrate it of their own accord, without a partner's prompting. Existing strategies that seek to modify children's environments in order to motivate their communication (see Chapter 2) have reported largely anecdotal, clinical success with little empirical support. The results of studies 1 and 2, particularly the latter, provide indirect empirical confirmation that the strategy of modifying familiar child environments to motivate communication *can* work. At least in a technology, this can be done fairly frequently and continues to work over a period of time. Section 9.5.4 comments at more length about how those existing strategies might be productively blended with the current design strategies.

In brief, additional contributions include:

New games A version of the *Andy's Garden* game suite was made available online in June 2015, playable through the free GameSalad arcade. This release was accompanied by information for parents, explaining the game premise, suggested manner/context of play, and troubleshooting actions using the keyboard hot keys. Game information and links were distributed to the families of all children in Study 2. As of spring 2016, all games are still available via the following URL: http://arcade.gamesalad.com/profile/AlyssaInformatics

New video dataset Video datasets of child-technology interaction are enormously rich, and can be analysed from many different angles. Given the difficulty in collecting such datasets and the high demands for access to children with ASC and special schools, a new dataset is valuable. The study 2 videos could easily be the basis of multiple undergraduate or MSc projects. These might focus on the effects of the current design, generating insights for future design, child communication, adult support and interaction, or even the difficulties with the touchscreen.

9.5.2 Technologies as tools to investigate autism

Most existing ASC-tech projects can be considered "tools for doing". They provide support in daily tasks (e.g. vSked visual schedule; Yeganyan et al., 2010; Hayes et al., 2010), teach academic skills (for reviews, see Pennington, 2010; Knight et al., 2013), teach social and language skills (for a review see Ploog et al., 2012), and can support parents and teachers in creating materials and strategies (e.g. a social story authoring tool, Constantin, 2015). Sometimes, technologies can be tools for sensory play, relaxation, and enjoyment (e.g. ReacTickles; Keay-Bright, 2007). Studying children with autism using research technologies, such as ECHOES and *Andy's Garden*, has an obvious goal of improving next-generation research technologies and moving toward publicly available tools. By learning from child-system interaction, we can make technologies easier or more interesting to use, and better at whatever it is they try to do. However, it is also possible to view this work from a very different angle: one in which autism and technology research is not only about researching technologies for people with autism, but about using technologies as a way *to research and understand the autism spectrum itself*. They can be "tools for knowing" as well as tools for doing. This distinction is about seeing technologies as tools or methods by which to conduct research, not only as an "end product" of research, or as a self-contained sub-field (where work with technology only comes from, and can only inform, other work with technology). Pieces of research, such as the current one, may play a dual role: producing data about specific technologies, and also providing a novel angle from which to consider autism. Useful observations of the latter types may even emerge naturally from studies primarily interested in something else. Hourcade et al. (2011) touch on this point in their discussion of tablet apps, noting that "seeing the different patterns children with ASD follow when using technologies as compared with typically developing children can help us learn about how they process information, and how they see the world", giving an example of how children with autism approached drawing in an opposite way to TD peers: from details on out to larger elements (p. 165).

Many of the qualities that make technologies a useful means for achieving support goals also make them appropriate for deliberately investigating the needs and characteristics of the autism spectrum. For example, there is the ability to closely customise and repeat children's experiences, provide timely cues and feedback, and the so-called "cool factor" of devices that make children want to use them (or at very least mean that using them is unexceptional, and does not single children out compared to their peers). Perhaps most importantly, they have the potential for creating situations and experiences that might be too overwhelming or otherwise "not safe" in face-to-face situations[17]. Technologies may have affordances that other data collection methods do not.

In the present research, the apparent conflict between the autism literature's representation of the *need for sameness* and child interactions with technology has already led to the development of one useful working hypothesis about how sufficient sameness may be able to "balance out" changes and rule-violation. This suggests a useful insight for how the need for sameness (or rather, the challenges around *lack* of sameness) may sometimes operate in face-to-face contexts. This prediction is definitely worthy of future investigation, and technology seems like by far the most ethical way to conduct early-stage investigations to refine the concept of balance, because it may feel safer or more contained. This in itself suggests a possibility for future intervention, discussed below.

9.5.3 Technologies as mediators for interaction

The current work contributes to the ASC-tech literature—and vast mire of commentary on children's technology use— by providing a clear counter-example to the reasonable fears that technology use might further isolate children with ASC and detract from their interactions with others. The original ECHOES system and *Andy's Garden* represent contexts in which a technology is frequently a *mediator* between a young child with autism and an adult social partner. It can be an attractive object of their joint attention, and a subject of communication[18]. Its internal pacing and structure can help create gaps for communication, as well as interesting things to communicate *about*. Dis-

[17] Parsons et al. (2006) give a good example of this, and report adolescents with autism trying out different behaviours within virtual environments—sometimes deliberately trying to see what would happen if they broke the rules. They note these explorations were sometimes also a good basis for discussion.

[18] Some of these ideas are discussed at more length in relation to the original ECHOES (Frauenberger et al., 2013).

crepancies, including child encounters with inherently novel game elements, seemed partially successful at getting children to turn to the adult to make comments, seek information, and share affect. Even the touch-screen presented a level of desirable difficulty, and encouraged children to ask for help and explain their desired actions within the game.

From the video data, it is extremely clear that children were spontaneously and easily shifting between interaction in the digital world and interaction in the "real" world, sometimes engaging in concentrated social interaction, and sometimes having longer periods of focus on the technology. They do not need to interact with either one all the time. There may be a much lighter social demand on children because they *can* switch in and out of communication with an adult, and do not have a constantly present demand.

In future work, there is obvious capacity to build on this role of technology as a mediator for social interaction by changing the way that *social partners* think about their own role: they are one part of a "triangular" social interaction with the child and a technology at the other points. They may help children to use the hardware (e.g. as with the current touch-screen troubles) or to understand instructions, but the technology can also help *them* support communication in a very real way.

9.5.4 Insights for face-to-face autism support

9.5.4.1 Creating communicative opportunities in physical settings

Looking outward and ahead, the insight about sameness apparently balancing unpredictability in virtual contexts may have applications for using technologies as mediators for (or objects of) communication, and may perhaps represent a way to transition to using skills in other settings. If the working hypotheses in this thesis truly tell us something *general* about the need for sameness, motivation, and the autism spectrum (rather than something only applicable to the current contexts, or to virtual contexts), the principles derived from and initially tested via technologies could form the basis of "face-to-face" activities or intervention.

The clearest example of this is to marry existing strategies about communicative "demands" in children's environments with the current strategies for embedding attractive, ambiguous DDOs in virtual settings. It is perfectly plausible to implement DDO-type situations in a child's physical surroundings: situations that alter a familiar or environment or introduce a novel element, but are not announced as a problem to solve and—the crucial difference from existing strategies—do not require a *specific* communication. For example, instead of giving the child his usual yoghurt pot but a too-large spoon that won't fit (Layton and Watson, 1995), the child might get the normal yoghurt with spoon, served in an unfamiliar container. This is an equivalent of a DDOs that changes something noticeable, but does *not* does not block an activity goal. Here, it is now up to the child to communicate about his daily yoghurt being served in the bottom of a giant mixing bowl. Perhaps he exclaims in surprise, asks for the usual container, laughs at his silly parent, or makes no comment and just eats the yoghurt. However, it *is* up to him, and this remains an *opportunity*, rather than a demand. If implementing these types of DDOs in physical settings, it will also be important to consider the high-level principles: for example, planning how to *resolve* opportunities, where relevant, or how to maintain the *integrity* of activities and goals. As in a technology—and perhaps even more important in the physical environment—there is an imperative to choose DDOs that will not make a child's environment appear lawless or create confusion about what the current activity is.

Ambiguous, DDO-type opportunities also complement face-to-face "demand" strategies because they are less prescriptive (or not prescriptive) with respect to child behaviours. This opens up possibilities for more socially-motivated behaviour (e.g. commenting, sharing information or affect). The clinical anecdotes and limited empirical findings described in the literature (see Section 2.4.2.1) all have created communicative demands answerable with specific, imperative communications: Help me, stop that, give me that. They are all about gaining access to objects or controlling the environment, including other people. Creating opportunities for socially-motivated communication—or at the least open, ambiguous opportunities—would be a valuable technique to extend the current strategies, especially for children who may already use imperative communications.

9.5.4.2 Desensitisation to discrepancy

Previous research on interactive virtual environments and games has explored their use as a way to prepare people with autism for real-life situations (see Parsons and Cobb, 2011, for a summary of recent evidence and issues). The core idea is that a technology-based experience can capture features of the real experience, simulate it, and help people become prepared for and comfortable with it. This allows us to speculate that games and programs with DDOs may be able to play a similar role. If children repeatedly experience discrepancies in a virtual environment, communicate about these, and even find them humorous, it raises the possibility that the technology might have a de-sensitising effect. Or, viewed another way, it might help to lessen the need for sameness. Just in the way that a VE might make a bus less scary, a game with surprises might make real surprises less scary—especially if children are *also* having the experience that changes and differences can be worked around, or resolved. Alternately, technology could be combined with face-to-face approaches. Experiencing discrepancy—within the confines of a game, or a device—might provide a transition into slowly introducing more changes, novelty, or surprises into a child's real environment. In some cases, it might allow a child to "try out" changes virtually first, in increasingly unpredictable scenarios.

9.6 Future work

9.6.1 Further analysis: function of child initiations

The present analyses have treated initiation as a general category of behaviours, set in opposition to non-socially-directed reactions, and has not coded initiation behaviours or functions. Analysing initiations in more detail can be carried out with the existing datasets in studies 1 and 2. This both provides greater detail of child behaviour within DR pairs, and offers the possibility of determining how types of initiations might relate to discrepancy subtypes or discrepant aspects. Here, annotating initiation function should be a priority because of the developmental benefit of certain initiation types, which future designs may wish to particularly promote.

The autism literature indicates that not all types of initiations are equal, in terms of how prevalent they are and their implications for development. As noted in the literature review, a key distinction is between *imperative initiations* that function to gain access to objects or otherwise manage the environment, versus *social initiations* for sharing affect, information, or social actions with others. In natural settings, children with ASC are observed to make far more imperative initiations (e.g. Chiang, 2009; Stone and Caro-Martinez, 1990). Thus, it is particularly interesting that a number of the DR pair examples in both datasets appear to show socially-motivated initiations. For example, the frequently-impassive Bradley (study 2) reacts to Andy's sudden disappearance by turning to make eye contact with the researcher and share his surprise (see Section 8.4.3). Or, there are the children who exclaimed and drew the researcher's attention to colour-changing flower pots and thunder sound effects in ECHOES (see

Table 6.1). All these child reactions appear to be pure social sharing, with no attempt to manage the environment, or gain access to any tangible objects. They don't "get" anything from initiating in this way, other than the social interaction itself. While some child initiations around ECHOES and *Andy's Garden* do seek to manage the environment or behaviour (e.g. ask for help), *none* are tangibly maintained by physical toys, snacks, or other motivators of the type common in the existing communicative "demand" literature (see Section 2.4).

Additional video annotation would be needed to produce a full picture of children's initiation functions in relation to discrepancy. If a sizeable portion of DR pairs appear to have a social function, this would be very exciting. It would suggest that detecting a discrepant aspect may be a strong extrinsic motivator for social interaction. Indeed, it may be a more effective motivator for social behaviours than the withholding of an extrinsic reward (of snacks, food, a completed routine) if the child does *not* initiate—the strategy in existing communicative "demand" examples. Extrinsic motivation for social interaction—especially one that is more cognitive or experiential, and does not rely on tangible reinforcement—would be a valuable tool.

9.6.2 Further analysis: reciprocal interaction sequences

Both current studies distinguished between a child's *primary* and *secondary* (or first and subsequent) initiations concerning the same instance of a discrepant aspect. These usually cluster together closely in time, though occasionally represent a child returning to an earlier event or topic of conversation. Scrutinising multiple initiations in relation to partner behaviours (or lack thereof) would be very worthwhile, to look for clusters of initiations that represent *reciprocal interactions*. This is a useful concept borrowed from the SCERTS autism framework (Prizant et al., 2005) and refers to a multiple-turn interaction in which the child and the (same) social partner exchange two or more consecutive "pairs" of contingent actions. These do *not* need to be linguistic, and could use other social actions or a combination of language and actions. At least one of these pairs must include an initiation from the child; there is no requirement for initiation from the social partner. Reciprocity is more complex than "one-off" initiations, or single child-partner exchanges. An ongoing interaction requires social partners to coordinate their efforts, if they are to make contingent, relevant interactional moves. Contingency (especially in verbal language) is thought to build on simpler joint attention skills because of this coordination requirement: it requires both parties to coordinate their communicative efforts. It is also required for more advanced social interactions and shared forms of activity (Tager-Flusberg and Anderson, 1991). Future iterations of design guidance might choose to prioritise fostering reciprocal communications.

While it is positive that discrepancies motivate *any* initiations, it would be even more positive if they could regularly motivate children to maintain contingent interaction over several turns, giving the social partner additional opportunities to scaffold the interaction. The descriptive data from both studies suggests that discrepancies do motivate reciprocal sequences for some children, but this needs to be systematically analysed and also considered in light of available information about individual children's language and level of development. In an analysis and in revised design guidance, partner behaviours should not be neglected. While having a general role in mind (e.g. to be reactive, not to "teach" the child about the environment) is a useful level of guidance, offering explicit goals or techniques for extending interaction might help to make the most of children's spontaneous initiations. Existing work on facilitation of contingent dialogue in autism may offer additional concrete suggestions (Tartaro and Cassell, 2008; Curcio and Paccia, 1987).

9.6.3 Refining the concept of "sameness"

The current design principles and design context assumptions (Chapter 6) are highly concerned with *sameness* as a way to balance the novelty and alteration of DDOs. Broadly, sameness refers to elements that are held constant within the design and its context of use. In *Andy's Garden*, this included things like Andy's hello/goodbye scenes, menus, and the location and duration of sessions. Sameness within a technology and in its context of use have been assumed to *both* play a role in a motivating but emotionally manageable experience, but this hypothesis has not been explicitly tested. Future work should try to tease apart the effects of both types of sameness, to learn what they contribute individually, and if one might be more important than the other in making discrepancy manageable. It is perfectly possible that in-game sameness is enough, and context actually contributes very little. Or, perhaps children can tolerate any level of discrepancy and change as long as it is contained within the technology, and is surrounded by a familiar setting and supported by a familiar adult.

The main challenge to investigating sameness is doing it in such a way that children feel safe and supported during technology use. For example, a design that dramatically reduces in-game sameness, should definitely have a familiar and supportive adult standing by during use. This type of investigation will probably need to be undertaken in steps, feeling out the effect of one type of change before trying another. Always, the designer should consider the overall effect of the design, and how children might experience it. This type of information about sameness would be particularly important for designers interested in use cases where sameness of context cannot be guaranteed, in terms of people present, physical location, length of play, and so on. For example, designs for mobile devices are much more likely to be used in variable contexts, simply because they easily *can*. Evidence from the current studies does not allow us to extrapolate about that type of use; new work is needed.

Further investigation of sameness and its contribution to child experiences would feed into updating the design guidelines initially proposed in this work. The greatest impact would likely be on *integrity* of environments, and *frequency* of DDOs in relation to other content. More information about the nature and role of sameness would also allow an update of the assumed contexts of use, adjusting or even substantially relaxing the “requirements” for sameness in physical environments and sessions.

9.6.4 “Going bigger” with DDOs in new designs

New designs that investigate more, “bigger”, and different DDOs is a natural extension to the current work. With more data, certain types of DDOs may begin to emerge as better than others for motivating certain communications, or for generally motivating certain sub-groups of children to engage with the system. The current design chose its DDOs cautiously, paying the utmost attention to integrity, and not upsetting children. In development, Andy’s end-of-activity disappearance was nearly cut out as “too upsetting” or “too confusing”. In evaluation, children greeted this event with hilarity, surprise, or utter indifference, and took for granted that Andy would reappear as usual at the start of the next activity.

Children’s interactions during the evaluation study suggest—as already discussed at the end of Chapter 8—that DDOs could “go bigger” and still be balanced by a similar amount of sameness as in the current games. There are several possible areas of investigation, which could all be explored in a single design to some degree, or more systematically investigated in different designs. Exploring them in a mixed design might be a better choice, due to the variety principle. Possible ways to make DDOs “bigger” might include:

- More attention-grabbing sensory properties: colour, motion, size of added or altered objects, animations, sound types and volume, etc.

- Changing appearance or effects of more central elements (e.g. could change Andy's appearance, rather than the background appearance).

- More incongruous DDOs: introducing objects or behaviours that are not closely related to the rest of the game setting and/or do not make sense there.

- DDOs that affect goal-relevant objects or actions, but without seriously violating integrity (i.e. creating confusion about what the game goal is). The effects could be temporary. This was explored in a very limited way in *Andy's Garden*, for example with the missing and re-entering cloud, which was fairly successful as a motivator.

- DDOs that affect cause-and-effect relationships or game mechanics, rather than "surface" elements.

DDOs that affect cause-and-effect relationships, activity goals, or game mechanics pose challenges related to resolvability, ambiguity of opportunities (rather than demands for action), and integrity. Investigating DDOs of this type will require a designer to either be very inventive, or to compromise on some of the principles. This is in fact a good way to test the amount of leeway in the current design principles, or alternate ways in which they could be implemented. For example, *integrity* might not need to be as absolute as currently thought, once initial expectations are established. Or, certain changes could be resolved through a researcher GUI, rather than by direct child interaction, timers, or self-resolution. Of course, child interactions might demonstrate that DDOs that change game mechanics (etc.) *are*, in fact, overly confusing and disruptive for the current target group.

9.6.5 Longer-term investigations

The previous suggestions for future work are all proximate steps, directly extending or building on the current studies. However, it is also possible to look at the current work as the first step on a longer arc of investigation, working toward behavioural interventions that *improve* or *increase the use of* children's initiation skills. These might be technologies, or new face-to-face strategies using insights from ASC-tech studies. These future intervention tools may primarily rely on discrepancy-detection as their mechanism of change, or might use it as one design strategy among several.

The thesis has currently framed much of its activity in terms of transferability, and applying insights from one context to another in order to create specific patterns of interaction. It is also possible to look at these activities and associated research questions in terms of *efficacy*, or the capacity to produce effects, asking "does the strategy of deliberate DDOs work to motivate initiation, under relatively prescribed or 'ideal' circumstances?"[19]. Rao et al. (2008), Lord et al. (2005), and Smith et al. (2007) all seem to agree that single case (and, by extension, small-n) research is a good way to establish efficacy. Smith et al. (2007) point to peer-mediated social skills training (e.g. Strain, 1977) and video modelling (e.g. Charlop and Milstein, 1989) as now-established autism intervention methods that have been "built up" from initial small efficacy studies.

If taking a long-term view of future work, the current studies are potentially the first part of the small-n, efficacy stage, in a longer path of steps toward validation, heading toward an initiation intervention strategy with real-world effectiveness. The current studies are definitely not *efficacy trials,* which is to say closely-controlled trials with strict inclusion criteria, strictly defined outcomes, and closely-monitored delivery (Lord et al., 2005). They are asking an earlier question: can we create this effect (initiation) *here,* during this game?

9.7 Summary of key findings and conclusions

1. In video data from the ECHOES virtual environment, young children with ASC were observed to notice and initiate about novel or unexpected events that were were frequently caused by system errors, rather than being part of the system design. For example, they commented to the adult researcher about a newly-discovered object action, or the virtual character's activity mistakes. These observations were unusual, given the well-documented *need for sameness* in autism that would suggest children would find these occurrences disruptive, rather than positive and interesting. As spontaneous initiation is developmentally important and difficult to target in existing interventions, these occurrences were worthy of systematic investigation to try to determine what was happening, and why children might be reacting positively and socially to non-sameness in this context.

2. Iterative coding and categorisation of the ECHOES video dataset was used to

[19]To be clear, efficacy asks about something working to produce a result *at all*, rather than the degree to which something can produce a particular result (effectiveness).

describe this observed phenomenon and children's reactions to it. This process developed the core concept of *discrepancy*, or subjectively-recognised inconsistency between the environment and an individual child's body of knowledge and expectations. Discrepancy is a state of affairs rather than a discrete thing. It has two subtypes (novelty and surprise), representing different types of inconsistency. Due to the subjective and individual nature of discrepancy, the relevant unit of analysis throughout this work is the *discrepancy-child reaction pair* (DR pair). These concepts are concretely realised in a *video annotation scheme*, which uses information about the state of the environment and specific child behaviours in order to label child reactions, and child-environment relations (discrepancy type).

3. Study 1 applied the final annotation scheme to a subset of ECHOES participant data (n=8, M=7, F=1; mean developmental age 5 years, 9 months). This yielded a consistent pattern of DR pairs across children, activities, and sessions, with a high proportion of discrepancy-contingent initiations to social partners, and few instances of negative affect. These findings suggested it would be worthwhile to try to re-create these positive, developmentally beneficial interaction patterns *on purpose* in a new design. The designer would introduce objectively new or different things into a familiar environment, trying to create motivating situations that children would find *subjectively* inconsistent. These engineered situations are *discrepancy-detection opportunities*.

4. Study 1 findings apparently conflict with the *need for sameness* that is a central characteristic of autism. The working hypothesis of *motivating but manageable* interaction suggests a resolution: that discrepancy and change can be not only emotionally manageable but can be very interesting to children, *where they are balanced by sufficient sameness within and around a technology.* These concepts of motivating manageability and balance provide a clear guiding principle for future designs.

5. Descriptive and quantitative data, practical "lessons learned", autism literature, and other sources were used as a basis for formulating six high-level design principles for creating a *motivating but manageable* technology that includes DDOs. These spell out more concretely what was hypothesised to have "gone right" in ECHOES, and summarise what another environment should do to re-create similar effects: establishing and maintaining *integrity* of the environment and

activities, ensuring *flexibility* and *resolvability* of child-system interactions, and offering wide *variety* of DDOs at an appropriate *frequency*. DDOs should pose ambiguous *opportunities* for communication, rather than demanding specific behaviours or communicative forms.

6. Study 2 evaluated a set of three mini-games with embedded DDOs (*Andy's Garden*) created based on the high-level design principles. This was a proof-of-concept study with 10 new children in a UK special school (M=8, F=2; mean developmental age 3 years 7 months). Results clearly indicate that the designed DDOs succeeded in attracting children's attention and motivated them to initiate communication, similar to the interaction pattern with "naturally occurring" discrepancies in ECHOES. Children's willingness to engage with *Andy's Garden* and very low levels of observed negative affect indicate that the game play experience was emotionally manageable and non-threatening.

7. Overall, the patterns of discrepancy-contingent interaction in Study 2 were highly similar to those of Study 1, both qualitatively and quantitatively. Based on these results, the design principles, context assumptions, and other "lessons learned" have been successful in supporting transfer between the original ECHOES context, to a new context (*Andy's Garden*). These findings suggest that deliberately including DDOs in technology designs can be an effective way of motivating initiations to a social partner during game play. The work reported in this thesis establishes a foundation for future work investigating the potential for DDOs to motivate communications in non-technological contexts, or as a part of a behavioural intervention.

Appendix A

Description of learning activities in the ECHOES system

Child participants in the ECHOES summative evaluation study played a series of game-like learning activities. Early sessions with ECHOES focused on simpler activities (in the sense of their goals and required child actions) in order to familiarise the child with the environment and allow him or her to feel successful and become confident in using the touch screen. Later sessions introduced more complex activities, some of them requiring more precise touch-screen actions, more steps to completion, and so forth. In most sessions, a child might do several activities planned for that day, and then be allowed to choose between several options or request a favourite activity.

The following is meant to introduce the most commonly discussed ECHOES activities, relative to the analysis reported in this thesis, and to give a sense of how the major objects and the virtual character, Andy, behave. Note that Andy introduces each activity and models its main action(s) *every time* it is played, not only the first time. He only skips this if the child has begun manipulating objects (e.g has thrown a ball or grown a flower) before he enters, or while he enters. In this case, he reacts to the child immediately by saying “Good job” or “Wow” or whatever goal-specific phrase was programmed for that activity.

Cloud Raining These activities start in an empty garden, with the magic cloud above but no objects on the ground. Andy introduces both of these activities by drawing attention to the cloud and suggesting rain. He then demonstrates jumping and “shaking” the cloud in order to trigger a rain animation and sound effect. Where rain touches the ground, a flower will begin to grow. More rain produces a larger flower, up to a maximum size. Flowers can also be moved around the screen, and can be placed anywhere (i.e. do not have gravity). If there is a pause in the activity, Andy will take a turn and then indicate it is the child's turn. The researcher could also manually trigger Andy's turn (or instructions to the child) through a GUI. In Cloud Raining, there is no set goal to be completed; there is no minimum or maximum number of flowers. A counter in the bottom left of the screen increments for each flower grown.

Flower Growing (in flower pots) This game is a slightly more structured version of Cloud Raining. There are five flower pots on the screen and the goal is now to grow flowers within the pots, rather than on the ground. There is a turn-taking element, as in Cloud Raining, though the child can play successfully even if ignoring Andy's cues. When triggered by AI planning or the researcher, Andy will physically indicate where he would like the child to grow a flower. However, the child can grow a flower anywhere. Andy says the activity is done when each pot has a flower, of any size.

Turning flowers into bubbles or balls This game begins with several flowers already grown on-screen, and no cloud or other objects. By using a "swiping" motion on the flower head, the child can transform each flower into something else. Swiping up turns it into a bubble which floats on screen, and swiping down turns it into a ball. The bubbles and balls have the same properties as when they appear in other activities. Once a child had grasped this activity, s/he could play it in only a few seconds by moving rapidly across the screen and swiping each flower. This activity turned out not to have good potential for supporting interaction, and held limited interest for children after the initial exposure.

Pot stacking The start of this activity looks like the start of Flower Growing, minus the cloud. There are five flower pots spaced across the screen, and the goal is to build them into a stack. Andy initially demonstrates picking up and stacking a flower pot. He will invite the child to take turns also, and can physically indicate which pot he wants the child to stack. When pots are stacked, they change colour and "stick" together (though can be unstacked, sometimes even by Andy due to mistakes in his planning). The game was finished when all pots were stacked into a single tower.

Throwing the ball through the cloud (Ball Throwing) At the start of the activity, a number of red bouncy balls fall into the screen from above and bounce around.

Andy enters and suggests throwing a ball through the cloud, which he then demonstrates. When a ball passes through the cloud, it changes colour, turning from red to blue, yellow, or green. Even if the same ball is thrown through the cloud multiple times, it never changes back to red. Andy always draws attention to the colour change, pointing at the ball and saying “Look, it changed colour!”. When Andy invites the child to take a turn, he indicates a specific ball via remote or contact pointing. However, the invitation and instruction tended to be ignored. A child’s action for “throwing” a ball was dragging it all the way to touch the cloud and then letting it drop; this was a source of some confusion because it did not directly follow on the VC’s modelled action. The child could also play with the balls by flicking or dropping them, or give a ball to Andy (which he would then throw). The giving convention in ECHOES is to hold an object over the VC’s body for several seconds, until it was accepted. The activity concluded when no red balls remained.

Sorting balls into boxes based on colour (Ball Sorting) At the start of the activity, nine bouncy balls fall into the screen and bounce around (three each in red, blue, and yellow). There are also three boxes on screen, one for each ball colour. Andy enters and suggests sorting the balls into the boxes, which he then demonstrates. Balls can only be sorted into a box of the correct colour; an incorrectly coloured ball rolls off the top. Once sorted, the child cannot remove a ball from the box, though the VC can (if he has made a planning error). When each box is filled it transforms into visual reward: red becomes bubbles, blue is a fireworks display, and yellow becomes several animated bees which buzz and fly around the screen before vanishing. When all balls have been sorted, the screen is empty and the activity is done.

Flower picking Three flowers are visible on the screen, at the far sides and in the centre. There is also a basket placed near the centre flower. Andy enters and introduces the purpose of the activity (picking the flowers into the basket), and asks for the child to give him the basket. Upon being given the basket, he would thank the child and put it down again. This initial part of the activity caused confusion due to its lack of discernible purpose. This activity was meant to be similar to the flower-picking task used in the ECHOES foundation study (Alcorn, 2010; Alcorn et al., 2011), in that the VC indicates one of the three flowers (through gaze, pointing, or both) and the child’s task is to touch the one that is indicated. When the correct flower is chosen, it flies to the basket with a fanfare. A counter in the bottom left of the screen increments for each flower picked. Unlike the task in the foundation study, the VC was planning his actions in real time instead of following a set script. The timing of the events was thus slow, erratic, and led to children poking at the flowers randomly while waiting for Andy to do something. Consequently, the activity was little used compared to the other options.

Tickle Andy In this activity there are no objects on screen, only Andy and a large tree which does not appear in any other activities. The tree's leaves can be pulled off, and will then "flap" like butterflies. They eventually lose energy and drift to the ground, but will fly again if they are shaken (in the same manner as the cloud). Andy does not give an introduction or any prompts. If he is touched on his torso, however, he will double over and exclaim "That tickles, stop it, that tickles!". As in cloud raining, there is no defined end point to this activity.

Appendix B

Behaviours that constitute initiation

The following table presents specific child behaviours that commonly form part of initiations, per the developmental psychology literature (e.g. Mundy et al., 2003; Prizant et al., 2005). These behaviours were used to assist in classifying child behaviours during the annotation processes in Studies 1 and 2. They are organised by the customary function of those behaviours, again per the literature. It only includes behaviours most relevant to the current context of interaction in ECHOES and *Andy's Garden*.

NB: The concept of initiation via *proximity* needed to be adjusted in light of how children used this behaviours in relation to the virtual character, during ECHOES sessions. Attracting attention via proximity often took the form of the child leaning in very close to the area of the screen where he was standing, and/or the child placing his or her face very close to the character's face.

Categories of initiation, by function		Qualifying behaviours
Social	Attempts to share affect, information, objects with partner; seek information; initiate social ritual or play. Behaviours in which the goal is interaction rather than being tangibly maintained by non-social reinforcers.	*At least one of the following:* Sharing emotion with a partner; Gaze shifting between a partner and a relevant object or screen area; Social referencing to a partner, i.e."sharing" gaze for the purpose of conveying affect/interest, or "checking" gaze for information about the current situation; Asking for information; Commenting; Trying to initiate a game; Showing an object (toward partner's face).
Imperative	Attempts to gain access to objects, or otherwise manage the environment and/or others' behaviour.	*At least one of the following:* Requesting an object; Requesting help or another partner behaviour; Protesting; Directing a partner to do something.
Attract attention	Deliberately attract social partner attention, with or without attempt to direct it	*At least one of the following:* Saying partner's name; Touching or pulling at partner; Using an object to attract attention (e.g. wave or shake object at, or near, partner's face); Proximity: approaching the partner and waiting.
Other	Behaviours that constitute initiation, but where their function must be determined from context (may be combined with behaviours in other categories)	Pointing and attention-directing gestures; Attracting attention to oneself prior to other communication.

Table B.1: Functions and constituent low-level behaviours of child initiation.

Appendix C

Annotation scheme manual

The manual reproduced in this appendix represents the final categorisation and rules of the annotation scheme, but does have some differences in *terminology* from the main part of the thesis. For example, it discusses discrepancies as "mismatches" between child and environment, rather than subjective inconsistencies. In the event of any apparent conflict, the text contained in the main body of the thesis represents the most recent and most authoritative version of the discrepancy concept and its subtypes.

The manual does not describe any of the design-specific annotation types that are reported in Chapter 8 (e.g. the "source" of discrepancies as DDOs, or system errors). It covers identification and labelling of discrepancy-reaction pairs only.

Taxonomising and Annotating Child Reactions to Discrepancy

Alyssa M. Alcorn (aalcorn@ed.ac.uk)

28th January 2014

University of Edinburgh, School of Informatics

Contents

List of Figures

List of Tables

Part I

Taxonomies of discrepancy and child reactions

1 Introduction and general principles

1.1 Organisation

This document is presented in three parts: The discrepancy and reaction taxonomies (I), notes about applying the taxonomies through a video annotation program (II), and a set of "worked examples" (III) which walk through several video segments step-by-step, explaining how the taxonomies were applied to these particular instances.

Part I begins with an introduction and then presents the discrepancy annotation scheme in two sections: One taxonomises and gives examples of types of discrepancies that a child may detect in the environment, and the other taxonomises and gives examples of a child's reactions to these discrepancies. The two halves are inextricably linked, as the current unit of analysis is the *discrepancy-reaction pair*, not "discrepancy" or "reaction" alone. Section 4 explicitly brings the two taxonomies together and explains how they work together during the process of annotating video data, and can be used to identify larger sequences of child and partner behaviours.

1.2 A child-centred view of discrepancies

Discrepancy is a high-level conceptual category referring to any instance in which a current aspect of the environment is mismatched with the child's current knowledge or expectations about the behaviour of the environment and its contents.[1] More concisely, discrepancies are mismatches between situation and expectation. There are several subcategories of discrepancy, with membership determined by whether or not the child has expectations about a given aspect, and the manner in which the aspect deviates from those expectations.

The current annotation scheme takes a child-centred rather than system-centred view of discrepancy. Discrepancy is not inherent to an environment, even through many examples of discrepancy may involve occurrences that are objectively "wrong" in relation to activity goals, maxims of relevance, or (though not applicable in the current context), factual content. Discrepancy is instead defined in relation to the *child's interaction with the environment* and the *child's understanding* of how that environment does or "should" work. Without a child using the system and making comparisons between a "mental model" of the environment and its current state, there is no possibility of detecting and reacting to a discrepancy. Thus, depending on what s/he understands about a particular environment, a child may not react to (or appear unsurprised by) events that are, objectively, in violation of that environment's established patterns, physical laws, etc. From his/her viewpoint, these aspects may be perfectly coherent and fulfil his/her expectations about how the environment should work. In this child-centred conception, there are thus no "missed opportunities" to detect and react to a discrepancy. Thus, the appropriate unit of analysis is the *discrepancy-reaction pair*: the child's reactions (or lack thereof) provide an opportunity for limited inference about whether s/he *had* any expectations about a specific aspect, and how they are discrepant from the actual state of the relevant environment.

As it is difficult for the adult researcher to take on the child's view of events, all definitions are given in terms of what a specific child appears to understand, know, or expect. The main evidence for this is the child's observable behaviour, and any statements the child might make about the environment or his/her

[1] The environment could be a virtual or physical environment. This document is mainly concerned with virtual contexts, but should also be applicable to physical contexts.

Figure 1: ECHOES experimental set-up, with a child participant at the main screen and the researcher seated at the control monitor (not visible).

own beliefs.[2] Of course, a young child's understanding of the environment is generally private, with explicit statements of expectation, belief, or prediction relatively rare. The main source of information available when attempting to infer the presence of discrepancy-reaction pairs is the annotator's knowledge of what the child has been exposed to in the environment (and how many times). What evidence might the child have about what is in the environment, and how it all works? In some cases, there may be additional information available from developmental psychology, such as guidance on what children of a certain [developmental] age are likely to understand about object permanence, cause-and-effect relationships, and other key concepts.

When there is no observable behaviour contingent on an environmental aspect (i.e. no reaction), there is insufficient information to determine if the child found *that particular aspect* discrepant, even if s/he reacted to other similar (or identical) aspects have elsewhere in the sample of data. If a child appears to be clearly surprised, and the researcher is not sure why he or she is reacting based on what is occurring in the environment at that time, then the child should still be the ultimate guide and a discrepancy-reaction pair should be noted.

1.3 Current context and data examples

The video data in referred to in the discrepancy-type and child reaction examples is single-camera digital footage of a young child using the ECHOES virtual environment, a cartoon-like "Magic Garden" in which various game-like activities offered opportunities to practice foundational social skills (Porayska-Pomsta et al., 2011; Foster et al., 2010). These were deliberately very simple with a focus on cause-and-effect relationships, due to the developmental ages of the target child users (late toddler to early school age). A virtual character, Andy, was both a guide and playmate in completing the activities. His behaviour was not pre-scripted, but deliberatively and reactively planned as the child played (for more details, see Bernardini and Porayska-Pomsta, 2013).

Children accessed the ECHOES environment through a free-standing 42" multi-touch screen. Children sat on a chair within easy reaching distance of the screen. During the summative evaluation sessions from which this data is drawn, an adult experimenter sat near (but not at) the main screen, controlling the environment from a second monitor (the screen of which was not visible to the child) and providing other support to the child (see Figure 1). As an ECHOES foundation study (Alcorn et al., 2011) discovered, children frequently interacted with the researcher(s) as well as the system. As a result, the camera was positioned to capture as much as possible of the broader study environment (screen content, child, and researcher).

[2] While the current document refers to young children who may not have strong language comprehension or production, there is no reason that the annotation scheme could not be applied to other users (who may be able to provide more explicit and detailed information about their expectations).

2 Taxonomy of discrepancy

The taxonomy begins by presenting some general terms used throughout the definitions of the discrepancy types (see Section 2.2). They are useful because they provide a level of abstraction above that of specific examples drawn from the data set with which the taxonomy was developed. The discrepancy types are then defined individually, with examples drawn from an existing video data set in which discrepancies and child reactions are present. The taxonomy of discrepancy concludes by presenting a decision tree for determining whether a child may detect some attended aspect in the environment as discrepant, and if so in what way. *There is no claim that the decision tree represents the actual steps in an individual's process of comparison when interacting with an environment and identifying a discrepancy.* Instead, the decision trees—like the rest of the taxonomy!—are meant to be a helpful tool for understanding the phenomenon of child communications motivated by discrepancy-detection.

2.1 General terms

2.1.1 Kinds and instances[3]

The *kind* is universal, rather than particular. A kind is a general type, and can be considered as both a "conceptual category" and as a set of features or constituent components that hold a certain relation to one another (i.e. exhibit a certain *structure*). One kind is discriminable from another because the instances of that kind are typified by a certain set of characteristics or properties *that are coherently ordered and/or related to one another in a certain way*, which is to say that the parts of the property set are not independently known, but are cohesive. While perhaps easiest to understand kinds of human-made physical objects, a kind might also be a species of plant or animal. Kinds also do not need to have any physical instance at all: they may be sounds with a certain relation (such as a musical composition), or may even be an idea.[4]

An *instance* is an individual example of a kind (i.e. singular in a physical, temporal, or other sense) that demonstrates the features and relations (i.e. the structure) that make it recognisable as being of that kind, or at least manifests *enough* of the characteristics that it can be identified as such, rather than as belonging to another kind.[5] Given that most kinds will be incompletely specified or otherwise "ideal", most kinds will thus have instances that vary in typicality, plus some "borderline" cases of membership. For example, it would be difficult to exactly specify the kind **CHAIR** or to identify an ideal instance *chair*. However, a range of examples are readily identifiable and discriminable from **TABLE** and **WINDOW**. It is also possible to think of "borderline" instances that share many typifying features of **CHAIR** (such as *bench, love seat,* and *bar stool*) that still seem to miss out on full membership. Determining kind structures and memberships will rarely be exact, and in many cases these determinations do not *need* to be exact in order to be useful.

Within the ECHOES environment (as introduced in section 1.3 and further discussed in Part III) **FLOWER** could be considered a kind [of digital object], and any individual *flower* would be an instance of that kind. All *flowers*, regardless of size, colour, or location on-screen, would still be instances within the kind **FLOWER** because they all have the characteristics that distinguish **FLOWER** from other kinds in the environment.

[3]The definition and discussion of kinds and instances draws heavily on Hausman's (1975) discussion of novelty, particularly Chapter 1, "Production and Radical Creation". Note that while I am indebted to Hausman, the terms "kind" and "structure" are used similarly, but not identically, here.

[4]See Hausman (1975, p. 23-25) for a discussion of some very specific examples of kinds, such as original paintings that are, arguably, both members of the general kind "painting" and simultaneously kinds for which they are the only *instances*. He implies–but does not discuss in depth– that the same argument may apply to an original musical composition.

[5]This is roughly analogous to objects in an object-oriented programming language (such as Java) being instances of classes, and inheriting the properties of their class (i.e. their kind).

> **Conventions for designating kinds and instances in-text:**
>
> **KIND** Henceforth each reference to a kind will be presented in bold capitals.
>
> *instance* Instances of kinds will be indicated by italics. Thus, aspect k as instance of kind **K**.

2.1.2 Kinds, instances, and judging novelty

Introducing the idea of instances and kinds becomes important for the current discussion of discrepancy when considering whether or not some aspect is *novel* (see Section 2.2.1). As Hausman points out, "each event or object in the world can be considered new with respect to its singularity", or its being particularly located in time and space, with other likely variation in its properties (1975, p. 20). However, this is in most cases not a very interesting or useful type of newness, even in the current analysis that explicitly concerns itself with what is new in the experience of particular individuals, rather than in all of human endeavour.[6] More interesting is newness of kind: encountering the first instance that evidences a heretofore unpredicted—even unpredictable—structure different from those already known; an instance that is not identifiable as a member of a known kind and is thus the first example of a new kind (i.e. it is *novel*, or has what Hausman calls "Novelty Proper" [1975, p. 28]). If some aspect is the first instance of a new kind, then the individual encountering it definitionally cannot have predicted it based on known kinds, and cannot have expectations about it. A new (singular) instance of a known kind is *potentially predictable*, even if it was not in fact predicted. The current analysis is concerned with this unpredicted newness of kind, rather than the newness of singularity. However, it uses the simpler term "novelty" rather than "Novelty Proper".[7] Novelty is further discussed in section 2.2.1.

2.1.3 Aspects [of the environment]

The term *aspect* is at an equivalent level of abstraction to that of *instance,* but points more clearly to the current context: individuals acting and encountering objects and events in a [virtual] environment. The following definitions of discrepancy and its subtypes refer to *aspects of the environment* (henceforth referred to as *aspects*) and their relationship to *kinds* and to other aspects. *Aspect* is an umbrella term that refers collectively to events, elements, non-events in an environment. Each of these terms is explained more fully below. While it would be possible simply to bypass the following terms and refer only to the more generic *aspects*, clearly defining the terms at this level (more general than specific ECHOES data set examples, but more concrete that instance/aspect) make the definitions easier to understand.[8]

Event: An *event* refers to an object behaviour, character behaviour, or audio/visual effect that happens in the environment, or that a user causes to happen in the environment through his/her interaction. In some cases *event* might describe a whole sequence of actions that generally occur as a unit.[9]

Element: An *element* refers to the contents of the relevant environment, and potentially the whole environment or context itself. This might include objects and social actors in the environment. In

[6] Hausman points to newness of kind as important and worthy of analysis while still declining to commit himself on the question of "novel to whom?". However, Boden (1990; 1995; 2010), who is interested in both newness to individuals and newness on a historical time scale, also draws an equivalent distinction between new (individual) things or ideas, and new *types* of things or ideas.

[7] The current annotation scheme also leaves some flexibility on what is considered potentially predictable: children are allowed multiple exposures to a novel aspect before the annotator assumes that they have formed some expectations about it, and that the structure of subsequent instances is predictable. See Section 6.2 for more on annotating novelty.

[8] While the current definitions refer to the context of a virtual environment (VE) and make reference to video data from the ECHOES project, they could be applied to another virtual context or potentially adapted for coding videos of a person interacting with a physical environment.

[9] An example from ECHOES would be treating Andy picking up a ball and putting it in the box as a single event, rather than three separate events of picking up, walking, and dropping.

the case of a virtual environment, an *element* refer to the whole environment and to its various digital contents, including any virtual characters. In both physical and virtual contexts, there is no requirement for elements to be things that the user can manipulate or otherwise affect.[10]

Non-event: The actual or perceived non-occurrence of an *event*. It may also refer to the actual or perceived absence or unresponsiveness of an *element*.

2.2 Subtypes of discrepancy

Discrepancy can be further sub-divided based on the type of mismatch between aspect k and the child's internal model of the environment. These relationship between these subcategories are summarised in Figure 2, and discussed in the following sections. Their core difference is that the novelty category *prohibits* expectations about the aspect in question, and the surprise category (and its subcategories) *require* them. Novelty and surprise are thus mutually exclusive in the current conception of discrepancy as a mismatch between situations and expectations. Currently, all discrepancies are sorted into one of these three subtypes (novelty, surprising events, and non-events), though additional categories may emerge during future analyses. The taxonomy would then need to be revised accordingly. Note that the "surprise" category has not been used directly in data coding; discrepancies have instead been labelled more specifically as surprising events or non-events (thus subsequent references to three types of discrepancy).

The section describing each discrepancy subtype includes a box with examples from the video data collected in the ECHOES project. See Section 5 for more about the ECHOES video data.

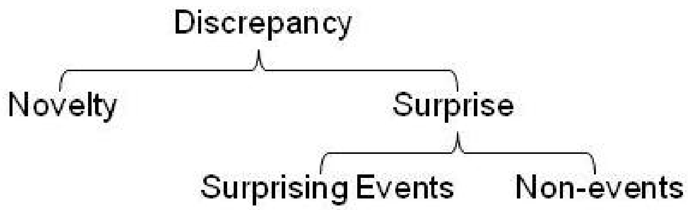


Figure 2: Diagram of the relation between the discrepancy categories and sub-categories. The "surprise" category has not been used directly in data coding; discrepancies have instead been labelled more specifically as surprising events or non-events.

2.2.1 Novel aspects (Novelty)

Novel aspects are those about which the child has not yet had an opportunity to develop expectations, such as one that is being encountered for the first time. The aspect is the first instance k of a new kind $\mathbf{K}$, which has a structure sufficiently different from that of other known kinds that it could not be reasonably predicted on the basis of experience with those kinds.

It is currently unclear how much experience (i.e. how many exposures) are required before an event or element should be considered *familiar* (see 2.2.2) rather than *novel*. This is to say it is unclear how quickly children should be assumed to develop expectations. The current annotation guidance suggests that three exposures should be used as an arbitrary cut-off, unless certain other circumstances are in evidence (see Section 6.2).

[10]In ECHOES, some elements are not manipulable and have no possible behaviours/actions, like the background artwork, the ambient garden sounds, and a minority of objects (e.g. the tree [trunk] in the "Tickle Andy" game).

Box 1: Examples of novel events (novelty) in ECHOES video data [that resulted in child reactions]

1. *Novel character behaviour:* The child plays the ball throwing activity for the first time, and sees Andy demonstrate throwing the balls through the cloud to change their colour.
2. *Novel audio effect:* The child successfully picks the flower Andy indicates, and gets the "music" reward for the first time.

For purposes of clarity and completeness, the scheme also defines the opposite of novelty, which is *familiarity.* This concept is never annotated explicitly, but is implied in the definitions of both surprising and expectation-fulfilling events.

2.2.2 *Familiarity

Familiar aspects are those about which the child has already developed expectations, though these need not be complete, nor completely correct. They are instances of a known kind **K.** This category includes non-events (i.e. there are expectations about aspects' non-responsiveness, absence, or similar).

As noted in section 2.2.1, it is currently unclear how much experience (or how many instances of exposure) are required before an event or element should be considered *familiar* rather than *novel,* though the suggested guidance is that expectations should be assumed to be present after three exposures (see Section 6.2).

2.2.3 Surprise

Surprising events These discrepancies could be most simply summarised as "expectation-violating events". An instance k of a known type **K** (or sufficiently similar to known instances of **K** that there could be reasonable expectations about its behaviour) occurs or is present in the environment. However, it does not appear or behave as the user expected or predicted, based on patterns, behaviours, or constraints already established through direct interaction with particular elements in a particular setting (such as a VE). The current k is discrepant from the expected k, but does not constitute a new kind **K'**. Some surprises may originate from the user's *incorrect* beliefs and expectations about kinds, the environment, or the effect of his or her actions.

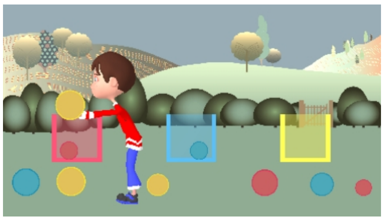
For some children, it may also be reasonable to discuss expectations based on the observation of others' interactions with the elements in that setting. Events may also violate expectations not specific to a particular setting, such as by violating physical laws of causal, spatial, or object relations that should be understood by a child of that developmental age (e.g. as discussed in Goswami, 1998).

Box 2: Examples of surprising events in ECHOES video data [that resulted in child reactions]

1. *Expectation about physical law not met:* The child manipulates a bouncy ball in such a way that it flies off the top of the screen and completely disappears, instead of falling back down again.

2. *Expectation about environmental rule/ characteristic not met:* In all activities except one, bouncy balls ricochet off the sides of the screen, though they may briefly go out of view at the top of the screen before falling again. However, in that activity balls can fly off the left and right sides of the screen instead of bouncing (due to a programmer error). After this had happened several times, children seemed to learn a new rule for the environment (created a new expectation).

3. *Virtual character mistake*: The virtual character Andy demonstrates a sorting activity, putting balls into the boxes of the same colour. After the child and Andy have both taken several turns, Andy then tries to put a yellow ball in the red box (see below). Balls of the wrong colour roll off the box's top, rather than falling inside.



[Surprising] Non-events These discrepancies are defined similarly to surprising events, except that they concern violations of expectation through aspects of the environment (instances k of known kind $\mathbf{K}$) unexpectedly/unpredictably being absent or failing to occur. In other cases, objects or social actors (such as a virtual character) may be present as expected, but do not perform any of their expected actions (based on their kind) or may not react to user actions that customarily produce a response.[11]

As with surprising events, some non-event discrepancies originate from the user's *incorrect* beliefs and expectations about kinds, the environment, or the effect of his or her actions. For example, while using ECHOES some children repeatedly requested help with "broken" objects that were not broken, but rather could not detect the child's inappropriate touch screen action (such as poking or scratching). From the child's view, however, there was a discrepancy between the expected result of the action and the object's failure to respond.

[11]In ECHOES, non-events do not mean a touch-screen failure or the entire system freezing; they are more selective issues with that affect some parts of the environment while the rest continues to function.

Box 3: Examples of non-events in ECHOES video data [that resulted in child reactions]

1. *Unresponsive element*: The child sits down to start a new session with ECHOES, touches his name repeatedly on the screen (the usual method of starting the first activity) and nothing happens.

2. *Unresponsive element*: In a flower-picking activity, Andy asks for the child's help and then gazes and points to one of the three available flowers. The child touches all flowers and none of them flies into the basket (indicating a correct choice). The child stops touching the screen and leans in to look very closely at Andy's face (i.e. social referencing; seeking information).

3. *Delay, perceived as absence:* The Magic Garden fades in to start a new activity. There is an unusually long pause without Andy entering (i.e. compared to previous activities), but the system does not appear frozen. The child asks, "Where's Andy?".

For purposes of clarity and completeness, the scheme also defines the opposite of a discrepancy, an *expectation-fulfilment.* These aspects are never annotated explicitly, as they do not provide any information about the goals of this analysis.

2.2.4 *Expectation fulfilment

The child encounters an instance k that has demonstrates the structure and behaviours that the child expected, based on his or her previous knowledge of kind $\mathbf{K}$. In some cases, expectations may be fulfilled by a lack of presence or occurrence (i.e. a non-event), such as when an element established as unresponsive does, indeed, not react to the child. Events (or feasibly non-events) may also fulfil expectations not specific to a single setting, such as when objects behave according to the physical laws of the real world and according to standard rules of causal, spatial, or object relations that are likely to be understood by a child of that developmental age (e.g. as discussed in Goswami, 1998).

Box 4: Examples of expectation-fulfilling events in ECHOES video data

1. *Consistent character behaviour:* The child has been playing the "tickle Andy" game. He tickles Andy again and Andy reacts in the way he has reacted to previous tickling, by laughing and saying "stop it, that tickles!"

2. *Consistent lack of responsiveness (i.e. expectation fulfilled by a non-event):* The child has established that the gate and hedges pictured in the Magic Garden "backdrop" do not respond when touched. The child touches them again and nothing happens.

3. *Object follows familiar physical law:* The child lifts a bouncy ball to the top of the screen and drops it. The ball falls and then bounces several times.

2.3 Decision tree: Discrepancy subtypes and expectation fulfilment

The relationship between the layers of discrepancy subtypes (and related categories such as expectation fulfilment) can be illustrated as a decision tree (Figure 2.3), using the language of kinds and aspects (as explained in Sections 2.1.1 and 2.1.3). *There is no claim that this diagram represents the actual steps in an individual's process of comparison when interacting with an environment and identifying a discrepancy.* Instead, it lays out a process of comparison as it may be followed in the annotation process, explicitly identifying the relevant criteria for determining the type of "mismatch" that appears to be present.

This diagram does not "stand alone" in illustrating the taxonomy. As discussed in Section 1.2, the unit of analysis is the *discrepancy-reaction pair*. Thus, in order to find a complete and analysable discrepancy-reaction pair, annotators must also consider the child's reaction (and, in some cases, its relation to the social partner's actions and prior child actions). Section 3.5 presents a decision tree for child reactions, and the summary in Section 4 unites both diagrams to show the complete process of determining whether an environmental aspect and child reaction together constitute a discrepancy-reaction pair, and may extend into a reciprocal interaction.

3 Taxonomy of child reactions

The taxonomy begins by presenting some general terms related to social communication which are drawn from the developmental psychology and autism literature and are used throughout the definitions of child reactions. They are useful because they provide a level of abstraction above that of specific examples, with specific types of behaviour or communicative content. The high-level reaction categories are then defined in individual sections. The taxonomy of child reactions concludes by presenting a decision tree in Section 3.5 by which it is possible to determine how a child reaction should be classified, dependent on previous child and/or social partner actions (if any).

It is very important to note that the annotation scheme discusses child *reactions* rather than child *responses* (except in a few special cases, as a part of reciprocal interactions) because the discrepant aspects that appeared in ECHOES are not socially directed. In other words, the child definitionally cannot be responding because the surprise itself was not a bid for interaction with the child (see the discussion of *response* in Section 3.3.1). However, for future annotation contexts this definition may need to be amended (or another reaction definition created) to include response, should the context include socially-directed discrepant aspects.

Non-social reactions to discrepancies are recorded in addition to initiations because these also help to fill out the picture of which events are perceived as surprising. A non-social reaction means that a child has definitely noticed an event, even if s/he has not initiated about it. As with non-reciprocal sequences of behaviours, they represent a likely *opportunity* for initiation if the environment of child, technology, and social partner could only be slightly modified in some way. They are thus of particular interest when considering design decisions about how best to facilitate communication via future technologies.

3.1 Initiation to a social partner (initiation)

When a child is *initiating*, he or she is introducing a socially-directed behaviour not contingent on the social partner's previous interactional moves, which is to say that the behaviour is not a *response* to a social partner's action (see Section 3.3.1).[12] In some cases (such as in the midst of an ongoing interaction sequence), the child may introduce a behaviour that is *related* to a social partner's previous behaviour (i.e. is still part of the same interaction, rather than beginning an unrelated topic or interaction), but is not contingent on (i.e. not directly dependent on) the previous behaviour's form or content. These, too, should be considered instances of the child *initiating* if the responsibility for continuing the interaction clearly lies with the child, not the social partner.[13]

[12] This is a "generic" definition of initiation, not one specific to discrepancy detection.
[13] See the first worked example in Section 10 for a demonstration of how these criteria can be met in practice.

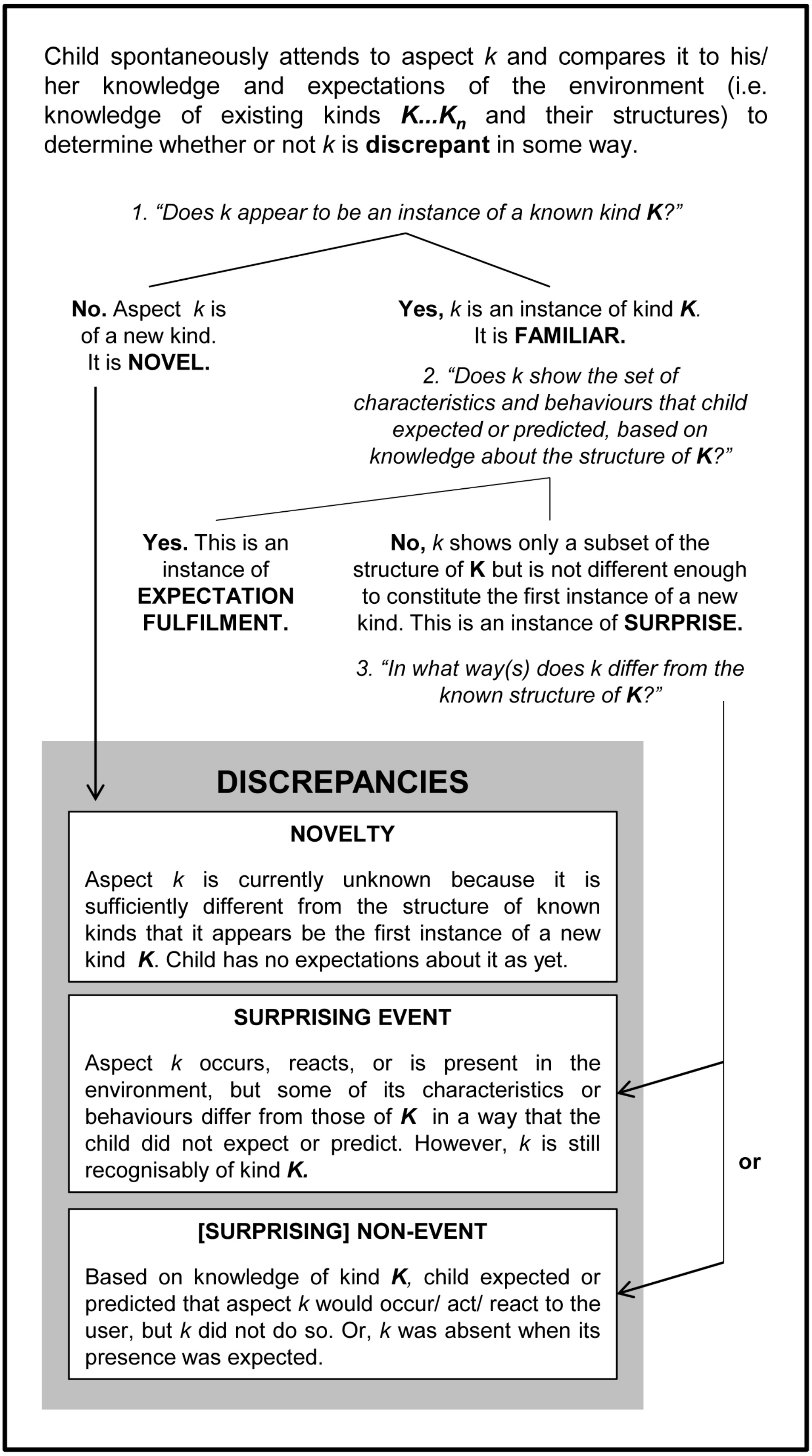


Figure 3: Decision tree for determining whether a child has detected a discrepancy, and what type of discrepancy it appears to be.

Initiation is a top-level category that includes behaviours with many forms and functions. An initiation behaviour may be imperative[14], part of an established routine such as greeting, or may be purely social, such as sharing positive affect. This project takes a deliberately broad view of initiation in order to best credit participants' communicative efforts, however unconventional those may be and whether or not they are acknowledged or understood by the social partner. Even a child with good productive language may make initiations where the specific goal or communicative content cannot be clearly determined except by someone highly familiar with that child. Unclear initiation goals or content are not obstacles to recognising child behaviours as initiative, as long as the behaviour in question does not appear to be contingent on the social partner's previous move, and fulfils three additional criteria[15]:

- Is the behaviour *spontaneous* (not the result of a prompt to perform that behaviour or a similar one)?
- Is the behaviour *purposive* (goal-directed), rather than accidental?
- Is the behaviour *directed* to a social partner?[16]

A child behaviour can count as an initiation even if the social partner does not appear to be aware of the child's effort, does not react to it, or apparently misinterprets its meaning. When it may be difficult if not impossible to determine what an initiation is trying to achieve, it simply does not make sense to "require" a partner response as a condition of a child making an initiation.

Child initiations can also be discussed in terms of whether they are *primary* or *secondary initiations.* These designations are concerned with the relation between the initiation and discrepancy, rather than the form or target of the initiation. These are defined in the following sections, and these categories' relationships are also mapped in Figures 3.5 and 6.

For purposes of clarity and completeness, the taxonomy of child reactions also includes the opposite of an initiation to a social partner, a *response to social partner* (defined in Section *3.3.1*). These reactions are only annotated explicitly as a part of secondary interactions, and may then form part of a *reciprocal interaction* about the discrepancy (see Section 3.4 for a discussion of sequences).

3.1.1 Primary initiation

A *primary initiation* is the first child initiation that is contingent on the occurrence of a particular discrepancy, or has content that is explicitly about that particular discrepancy. The initiation may be made to any social partner and make take any form or combination of forms. The initiation also need not occur immediately after the discrepant aspect is first encountered; it may be delayed.

The non-social counterpart to a *primary initiation* would be a *non-social reaction*; see Section 3.2.

3.1.2 Secondary initiation

If the child has already made an initiation contingent on the occurrence of a particular discrepancy, or with content that is explicitly about that particular discrepancy (a *primary initiation*) any additional initiations about that same discrepancy constitute *secondary initiations.* If, for example, the child sees Andy make a ball-sorting mistake and comments on it to the researcher, this would be a primary initiation. If

[14] These can be conceptualised as "tangibly maintained behaviours" (White et al., 2011, p. 1249) such as those that result in the child gaining access to an object, or bringing about another person's desirable behaviour. Imperative behaviours are often discussed as "opposite" to socially-motivated behaviours such as sharing affect, in which (at least for typically developing individuals) the goal is the social interaction itself.

[15] This definition of initiation is based closely on the one used in the SAP for ECHOES.

[16] Direction may be determined on the basis of content, physical direction to/ orientation towards the partner, or extrapolation from the child's usual patterns of behaviour.

the child then points to the correct box and gives an instruction to Andy, that would be a secondary initiation. Any following initiations about that same character mistake would also be secondary initiations. Were Andy to make another sorting mistake, the first child initiation about that specific mistake would then be a primary initiation, and so forth.

Some secondary, socially-directed reactions may occur following a relevant action from the social partner—leading to confusion over whether these are really initiations or should be noted as child responses. As noted under the main definition of initiation, in situations where the child's action may be *related* to the presence of a partner action but not directly dependent on its form or content, the behaviour should be considered a secondary initiation if the child is the one who takes the responsibility for continuing an interaction that would otherwise have ended.[17]

There is no non-social counterpart to a *secondary initiation,* because it would be unlikely that the child's behaviour could be confidently identified as being contingent on original discrepancy, rather than something else in the virtual environment.

3.2 Non-social reaction

When a child shows behaviour that is contingent on a discrepancy but not directed to a social partner (i.e. does not constitute an initiation), it is considered a *non-social reaction.* Note especially that verbalisations should not automatically be assumed to be socially directed; they may be self-directed.

The social counterpart to a *non-social reaction* would be a *primary initiation* (see 3.1).

3.3 Other types of relevant child behaviour

3.3.1 *Response to a social partner (response)

In order for the child's behaviour to be a *response*, it must be contingent on a previous social partner action that was **directed to the child**. In short, a child can only make a *response* if the partner has made an *initiation.* If the child's socially-directed behaviour is contingent on a proximate social partner behaviour which was *not* directed to the child, the child is actually *initiating* because the s/he is beginning a new interaction, rather than continuing a current one.

In addition to contingency on a partner's child-directed action, the general guidelines for whether a child's behaviour could be a response are summarised as follows[18]:

- Is the behaviour *purposive* (goal-directed), rather than accidental?
- Is the behaviour *directed* to a social partner?[19]

A *response* must also be temporally proximate enough to the partner behaviour for contingency to be apparent (though no specific time limit is set). As with the definition of *initiation* used in this scheme, a *response* does not need to be conventional in its form, and does not need to be correctly understood by the social partner.

The criteria for social partner responses [to the child] as mentioned in the dual-taxonomy flowchart (see Section 4 and Figure 6) are the same: purposive action directed to the child, and contingent on the child's previous socially-directed action.

[17]This issue is discussed as a part of worked example 1 and may be clearer in context; see particularly Section 10.5.2 and Table 3.

[18]This definition of a response is partly based on the one used in the SAP for ECHOES.

[19]As with initiation, direction may be determined on the basis of content, physical direction to/ orientation towards the partner, or extrapolation from the child's usual patterns of behaviour.

3.3.2 Prediction/expectation (a prediction)

A prediction is an utterance or other behaviour by which the child indicates that s/he has an expectation about an aspect of the environment, reveals the specific content of an expectation, or explicitly indicates a prediction about a future event or cause-effect relationship.[20] This may or may not be an event that the child has already seen. Predictions do not need to reflect actual states of affairs, nor do they need to meet the criteria for initiations.

Child behaviours of this type can help to illuminate what, exactly, the child understands about the aspects of the environment with which s/he is interacting and the events/non-events which s/he has witnessed or caused. Predictions can also help to answer the question of how soon the child may form expectations after initial exposure to novel events/elements.

Box 4: Examples of child predictions in ECHOES video data

1. *Expectation about a cause and effect relationship:* When all balls are sorted into the yellow box, the box disappears and cartoonish bees appear and buzz around the screen for several seconds. A child just putting the first ball into that box (on this iteration of the game; not his first time) says to himself "and bee goes dzzzzz!"

2. *Expectation about an object's behaviour:* When transitioning between ECHOES activities, a red bubble always appears in the centre of the screen and slowly floats off toward the top right corner. When the bubble appears on one particular occasion, the child points to the place where the bubble will exit the screen. The researcher asks him if that is where the bubble will go, and he says "yeah."

3. *Prediction about the rules of an activity:* On the first time playing the ball sorting activity (which Andy introduces by mentioning sorting, but not by explaining explicit rules), the child sees Andy pick up a red ball. The child spontaneously points to the correct box, and says "put a red ball in there." This is an example of a prediction (about the 'rules' of the activity) that is both correct, and meets the criteria for an initiation to Andy.

3.4 Sequences of interactions

Sequence Some discrepant aspects may motivate multiple child reactions contingent on that same event (or non-event). Any discrepancy followed by a primary initiation and then one or more secondary initiations is known as a *sequence.* The initiations may or may not be closely connected to one another by their content or proximity, and may or may not be interspersed with responses from one or more social partners.

Reciprocal interaction sequence Some sequences (or parts of sequences) may constitute *reciprocal interactions.* This is a concept adopted from the SCERTS framework (Prizant et al., 2005) and refers to a multiple-turn interaction in which the child and the [same] social partner exchange two or more

[20] It might be marginally more specific to refer to this category as "expectation" but that was deemed too confusing in light of how heavily this word is used elsewhere in the taxonomies. Thus, instances of this category will be generically known as *predictions*, until or unless the data suggests that there is reason to separately define and examine instances of child expectations and instances of child predictions.

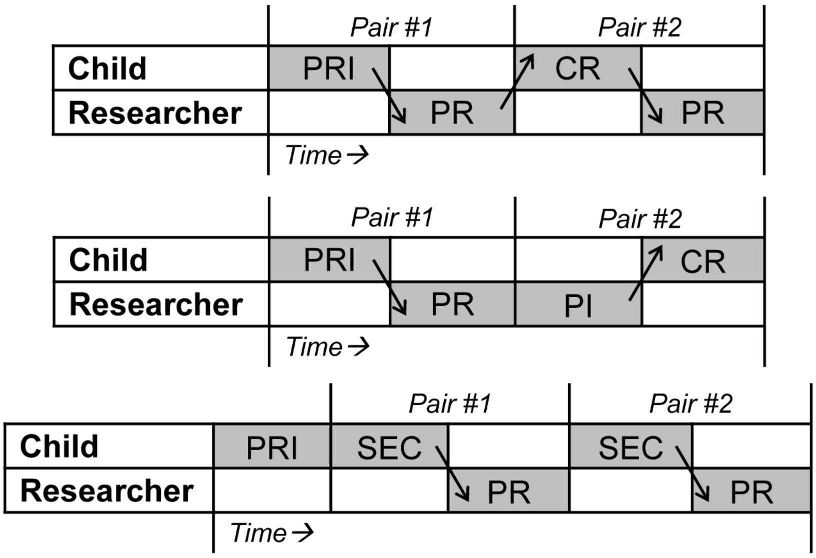


Figure 4: In all diagrams, arrows indicate contingency between "moves" in the sequence. *Top:* A simple reciprocal interaction in which the child's primary initiation is followed by a series of responses from both partners (i.e. a series of actions where each is directly dependent on the form and/or content of the partner's previous action). *Centre:* A reciprocal interaction in which one partner takes two actions in a row. *Bottom:* An example in which one partner makes multiple initiations before the other partner joins the interaction.

consecutive "pairs" of contingent actions. At least one of these pairs must include an initiation from the child. There is no requirement for initiation from the partner.

SCERTS subdivides this concept into *brief reciprocal interactions* and *extended reciprocal interactions.* The former require a minimum of two consecutive exchanges, and the latter require a minimum of four. Here, only the general term *reciprocal interaction* is applied to any qualifying sequence with two or more exchanges, regardless of how many exchanges there are.

There are multiple possibilities for how a sequence might qualify as a reciprocal interaction in the current annotation scheme, even given the indirect constraint that the first initiation [in the sequence] *must* be the child's. This constraint comes from the fact that—currently—the only type of social child reaction which "counts" in a discrepancy-reaction pair is *initiation,* which definitionally cannot be contingent on a social partner's previous interactional moves. While the child and partner *may* have a reciprocal interaction regarding discrepancy in which the partner initiates the first exchange (for example, after observing the child make a non-social reaction), the annotation scheme's current focus on initiation behaviour means that these sequences are of less interest and are not accommodated. However, the current annotation scheme easily accommodates instances in which the child may make *multiple* initiations, or even a non-social reaction followed by initiation(s), before the partner responds to the child and a sequence could begin. Several of the possible sequence types are illustrated visually in Figure 4, with arrows indicating the presence of contingency between interactional moves.[21]

The dual taxonomy flowchart in Section 4 and the "worked examples" of annotations in Part III may make the requirements for sequences and the types of sequences clearer: The former treats behaviours as states and shows how they can "legally" follow one another under the "rules" of the current taxonomies, and the latter give concrete descriptions of qualifying child and partner behaviours.

[21] These diagrams are adapted, with much thanks, from a style of "communication chain" notation presented by Jaegermann and Klein in *Enhancing mother interaction with toddlers*, presented at the International Meeting for Autism Research, May 2013, San Sebastian, Spain.

3.5 Decision tree: Child reactions

The relationship between the types of child reactions can be illustrated as a decision tree (Figure 3.5), using the language of kinds and aspects (as explained in Sections 2.1.1 and 2.1.3). In some cases, it is necessary to consider the child's own previous actions and the social partner's actions in order to determine what type of reaction is present.

As the diagram implies by its reference to *discrepancy-reaction pairs*, this decision tree does not provide complete information. Section 2.3 has already described a decision tree for environmental aspects, and the summary in Section 4 unites both diagrams to show the complete process of determining whether an environmental aspect and child reaction together constitute a discrepancy-reaction pair.

4 Summary: Linking the two taxonomies

When the two taxonomies are combined into a single flowchart (henceforth referred to as the *dual taxonomy flowchart*), it is possible to more clearly illustrate the idea of a discrepancy-reaction pair. This chart shows how the discrepancy decision tree (see Section 2.3 and Figure 2.3) and the child-reaction decision tree (see Section 3.5 and Figure 3.5) fit together, and how individual discrepancy reaction pairs—potentially in conjunction with other child and social partner actions—can build into larger sequences of behaviour. Some of these sequences can be considered reciprocal, per the definition in Section 3.4. The individual taxonomy flowcharts give the details for determining whether an aspect or behaviour "qualifies" as a member of a certain taxonomic category.

The current chart treats different actions as states, and shows all possible moves between states that are allowed by the category definitions. Where multiple links are possible from a given state, (e.g. from D-PRI pair to partner response and to SEC) this flowchart *does not* repeat the criteria for determining the next state, if indeed there *is* a next state contingent on the same original discrepancy. For example, to move from D-PRI pair to SEC, means that criteria for child action to "count" as SEC must be in effect.

Note that it is possible to exit the flowchart at any point, whenever it is not possible to make a "move" to any of the states that could legally follow the current state. In practice, this will often mean that the child has begun a new action that is no longer contingent on the original discrepancy (or on the current topic of the interaction, if the interaction is reciprocal and has evolved toward a new topic). Exits from this flowchart would also occur when the child or social partner stop responding to one another, or otherwise allow the interaction to lapse.

Another important point is that a child *can* complete a D-NSR pair and then immediately complete a D-PRI pair about the same discrepancy. The child could then go on to other social behaviours in the chart. However, it is *not* possible to move from the "social" part of the flowchart back to a NSR. The child may make other non-social reactions, but they are not counted as part of discrepancy-reaction pairs. This is because the annotation scheme currently has defined the rule that later non-social reactions (after the first) cannot be linked to the original discrepant aspect clearly enough to count them for analysis.

Partner initiations to the child require a special note. A "move" to the partner initiation state moves outside of the flowchart in the sense that partner initiations are not defined (or implied) in the child reaction taxonomy, but they still *may be* relevant to the child and to the evolving thread of the interaction [about a discrepant aspect]. Partners may, of course, initiate to the child at any point in time independently of what the child is or is not doing. The current ECHOES data set has many examples of the social partner initiating concurrently with the child, or initiating *instead* of responding to the child; this is particularly common when the child's verbalisation may be difficult to hear or understand. However, as in the definition of discrepancy, the current annotation scheme is child-focused and is primarily interested in *child initiations*. As a general rule the scheme thus prioritises the child's focus of attention

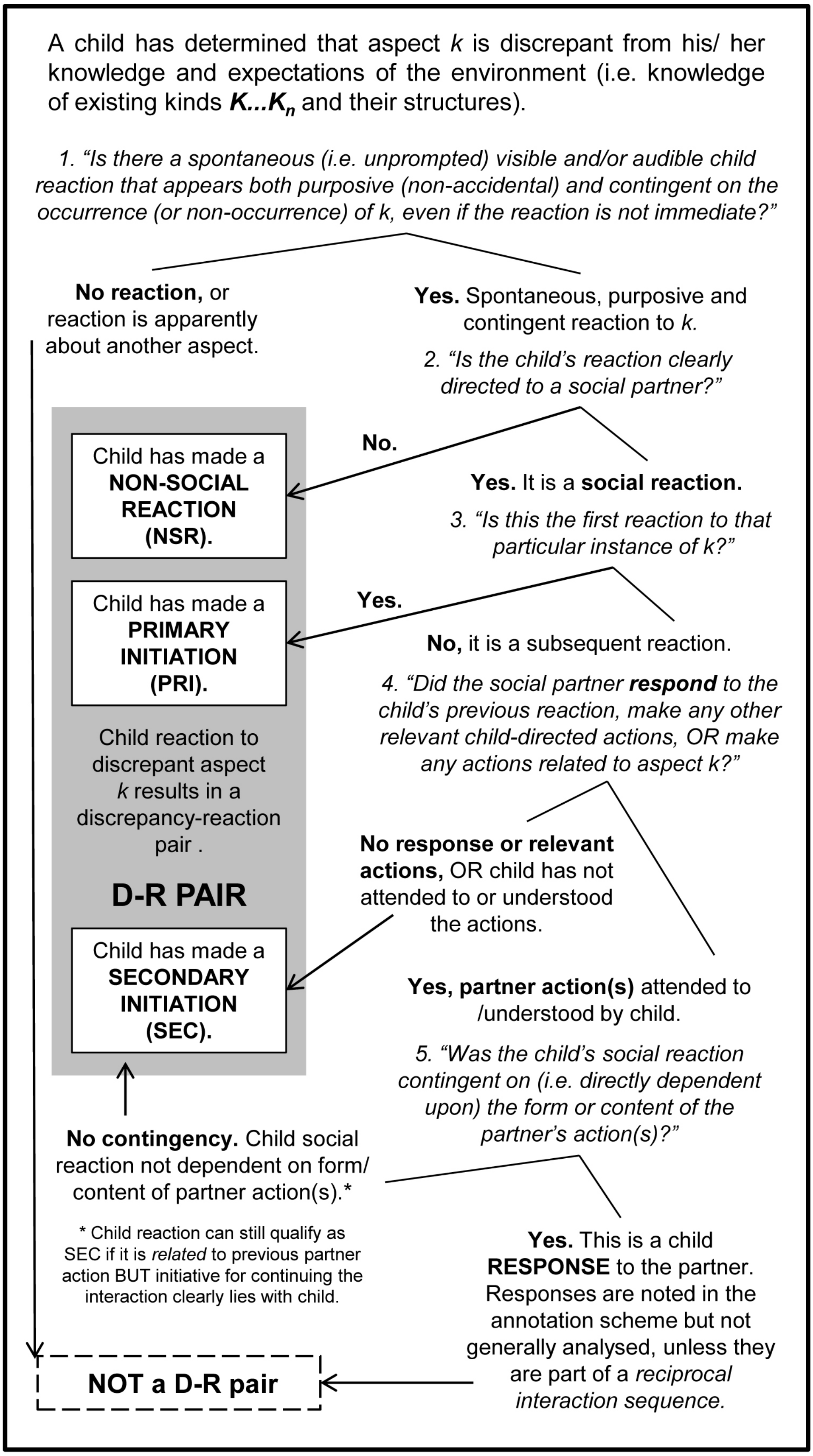


Figure 5: Decision for determining whether a child has reacted to a discrepancy, and what type of reaction it appears to be.

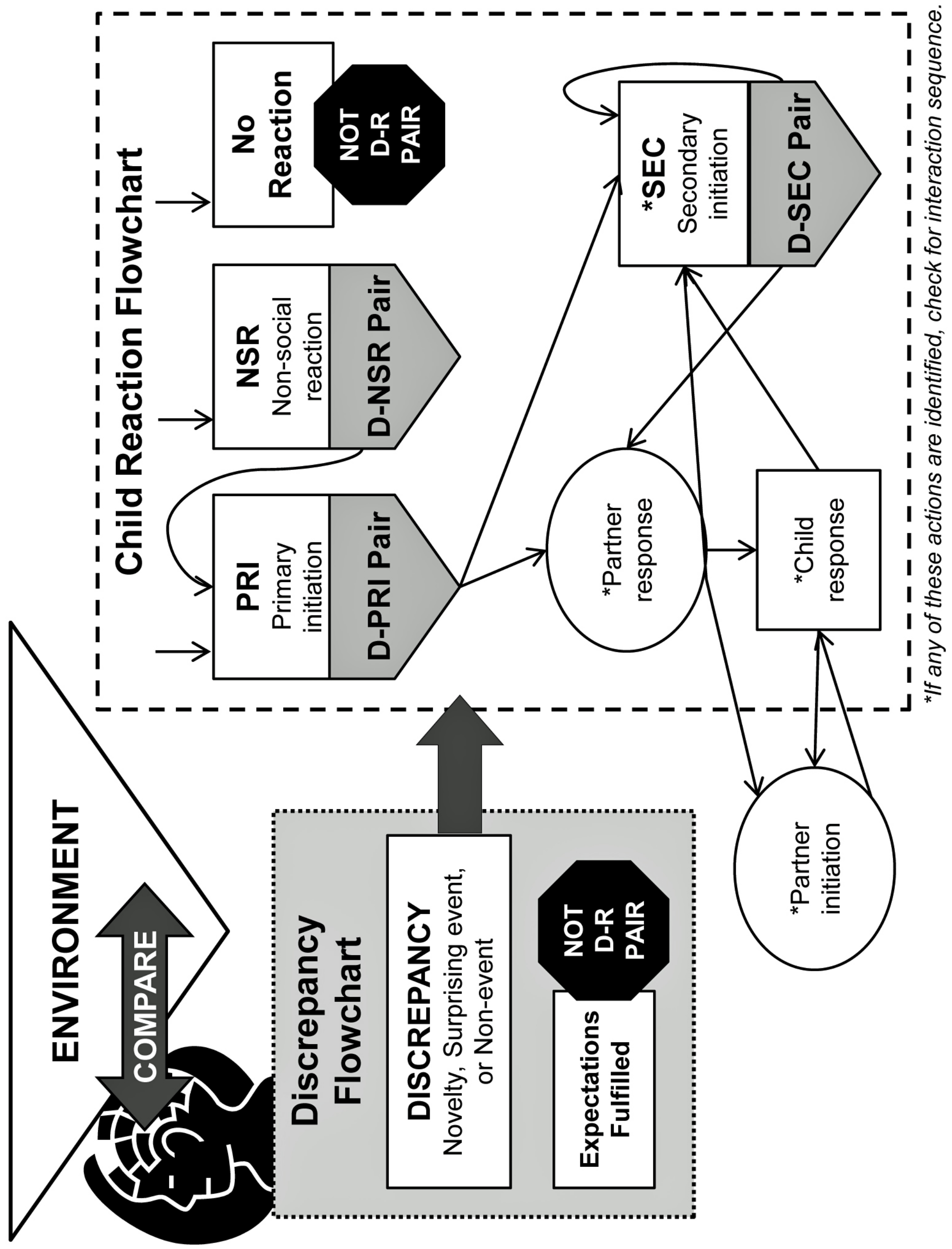


Figure 6: Applying the discrepancy and child-reaction decision trees in order identify valid discrepancy-reaction pairs and additional social actions that may contribute to reciprocal interaction sequences.

and action when determining “what is going on” and deciding what to annotate, only recording any partner actions where they could legally contribute to an emerging reciprocal interaction. This means that partner initiations would only be recorded following a partner response[22] or a child response, *when the partner acts to continue the existing interaction.* Thus there are limited links to and from the partner initiation state on the flowchart.

In summary, the partner's actions (even those directly related to the discrepant aspect) are not recorded *until and unless* the child appears to attend to those actions. This is already implied by the definitions of the various child actions: for example, PRI, NSR, and SEC all have as prerequisites that the child action is unprompted and not contingent on a partner action. It is important to remember that just because there *is* a partner action, this does not automatically mean that the child attended to it or understood it (and thus does not automatically mean that child actions are contingent on it). Worked examples 1 and 2 (see Sections 10 and 11) illustrate this point well. Both include concurrent, irrelevant researcher initiations which the children ignore (or do not even register) because they are pursuing communicative goals of their own. Thus, the subsequent child reactions qualify as initiations.

Part II

Using the taxonomies for video annotation

5 Overview

This part of the document begins with general guidance on how to consistently apply the taxonomic categories when annotating video data. It suggests “rules” for consistent application that are outside the scope of the term definitions and taxonomy flowcharts (Section 6). It then outlines how the taxonomies were applied as an annotation scheme to a corpus of video using the ELAN Linguistics Annotator, version 4.4.0[23] (Section 7). However, there are many available annotation programs and no reason why the taxonomies could not be applied through another program– they were not developed specifically for or within ELAN. Data can be extracted from ELAN as text, for manipulation in other programs, and thus this part concludes with a discussion of analysing the extracted data using a standard spreadsheet programme (Section 8).

As explained at more length in section 1.3, the video data referenced in the guidance examples is single-camera digital footage of a young child playing game-like activities in ECHOES with the virtual character Andy, using a touch screen. The child participants whose video data has been annotated for discrepancy-reaction pairs all have a previous autism-spectrum diagnosis. Thus, some notes on the annotation scheme refer specifically to autism-related challenges and concerns. Results of the specific analyses discussed as an example in this document have been reported in Alcorn et al. (2013).

6 Annotation guidance: Applying the taxonomies consistently

6.1 Locating discrepancies and reactions

The first step in annotating discrepancy-reaction pairs is identifying analysable episodes in the video segments. Taking a child-centred view of discrepancy notably complicates this issue, because there is no definitive list of “possible” discrepancies in a given piece of video (or in a system) against which a

[22] These have the same requirements as child responses. See Section 3.3.1 and Figure 3.5.
[23] This program is available free on-line from the Max Planck Institute for Psycholinguistics, at http://tla.mpi.nl/tools/tla-tools/elan/ (2012).

child's reactions may be checked. However, this certainly does not mean that children's perceptions of discrepancy are completely random or completely unintelligible to an outside observer: many *do* involve something objectively wrong, missing, or different, in comparison to the system's usual operation. Many more pairs include things that are objectively novel, not having been presented to the child before [in the context of ECHOES].

The discrepant [aspect]-reaction pair suggests two natural strategies for finding analysable episodes, further discussed in the following sections. The first begins with examining the aspects themselves, and the second begins with examining child reactions. Before describing either strategy, it is worth noting that the analysis does *not* proceed from the assumptions that children *only* react to discrepancy. Nor does it assume that the presence of a child reaction *necessarily* signals that s/he has detected a discrepancy (although it may signal this). Indeed, the majority of any child's reactions and/or communications about a particular activity, or to the VE as a whole, likely will not be motivated by discrepancies.

6.1.1 Strategy 1: Look for objectively new aspects and objective errors

In one sentence, this strategy involves looking for things that are objectively new or different than the currently known content of the environment (i.e. known *to the child* based on his/ her experience with the system*)*, and checking to see if the child appeared to react to them. The objective newness or difference is not important in itself: it is merely a useful "flag" for which parts of the dataset might merit more careful examination.

Some of the easiest discrepancy-reaction sequences to locate are those involving *novelty*, as defined in Section 2.2.1. In most cases, it is quite clear whether or not a particular child has previously been exposed to a given aspect[24] within his or her video sample.[25] In other words, there is a fairly straightforward judgement about whether or not something is "new". For more details on how *long* something was considered to be "new" to the child (i.e. child did not have assumptions about it), see Section 6.2.2. If the child reacts[26] to this new aspect, then this is an example of a [novel] discrepancy-reaction pair, and can be analysed further for its more detailed features.

After the annotator gains some familiarity with the "correct" functioning of the environment and the content of the video dataset, it is also fairly straightforward to locate things that are objectively different from the correct (or at least prevalent) behaviour of the environment (i.e. *surprises*). These might be actual *events*, such as Andy sorting a ball into a box of the wrong colour (as pictured in Box 2.2.3 and discussed in 11), or deviations in the form of object absences, event non-occurrences, or object/character unresponsiveness (i.e. *non-events*; examples in Box 2.2.3 and as discussed in 12). Many of these motivated child reactions of one form or another.

Keep in mind that it only makes sense to ask about novelty *or* errors in relation to any given aspect. To the annotator's knowledge, something might be both objectively new (to the child) and objectively erroneous (with respect to the customary system goals or function), but to the child it can only be one or the other. If the child has no expectations about a given aspect, how could she detect it as being erroneous or different? In the rare instance where a newly-introduced aspect may also involve an objective error, novelty will always be the overriding factor. The child then may react to an additional discrepancy (a surprise) on further presentations of that aspect, when it returns to its customary state (per the system design) but appears to be behaving in an expectation-violating fashion (to the child).

[24]May be an activity, digital object, sound effect, virtual character behaviour, etc. See section 2.1.3.

[25]These judgements may further be supported by system log files and researcher "session diaries" which listed which activities were played during which sessions.

[26]Per the definition of reaction in Section 3.

6.1.2 Strategy 2: Work backwards from child (re)actions

In one sentence, this strategy involves identifying video segments in which the child reacts to some aspect of the system, and then working backwards in time to determine whether that aspect may constitute a discrepancy. This does not mean investigating every single child action; some are clearly responses to the virtual character or a human partner (see Section 3.3.1 for the difference between responses and reactions). Many others are "business as usual", and involve the child repeating (or observing) favourite aspects/actions that are well known and show no notable deviations from customary patterns.

Types of child reactions that may be particularly worthy of investigation with respect to locating discrepancies are as follows (based on the current dataset and listed in no particular order):

1. Child appears surprised or puzzled about something s/he has observed in the environment (any valence)
2. Child appears surprised or puzzled about the result (or lack of result) from his or her own action (any valence)
3. Child indicates belief that something is broken, missing, or should be happening (but is not). In some cases this is very clear, as in the case of asking where something is, or declaring that an object is "stuck"
4. Child appears to find something humorous or otherwise worthy of exclamation
5. Child asks for help or information about something
6. Child appears to unsuccessfully attempt the same action multiple times in a row; child abruptly shifts attention after an unsuccessful action
7. Child appears to be watching the environment/ virtual character or waiting (without touching the screen) for an unusually long time
8. Child verbally contradicts the virtual character or verbally/non-verbally gives him a direct instruction
9. Child verbally or verbally attracts a partner's attention to something in the environment

Of course, there are many types of child reactions that may be motivated by discrepancies in addition to those suggested above, and not all instances of the actions above will be related to discrepancy. They are, like objective instances of novelty, simply "flags" for sections of video that merit being examined in more detail.

In some cases a child may clearly and repeatedly imply a belief that something is wrong or different about the environment, but to the adult annotator (or even the researcher in the video) everything appears to be following its usual pattern. It is important to repeat here that some surprises (both events and non-events) may originate from the child having *incorrect expectations or knowledge* about kinds, the environment, or the effect of his or her [physical] actions. For example, applying incorrect touch-screen actions (such as scratching or poking) led multiple children to believe that objects were sometimes "broken". The environment is operating correctly; the problem is with the child and his or her "model". Nevertheless, from the child's view there was a definite discrepancy between the expected result of the touch action and the object's failure to respond (a non-event), motivating exclamations about "stuck" or broken objects, requests for help from the researcher, and so forth. Another example would be the child who commented repeatedly and excitedly to the researcher about the "big flower" she had grown in one activity. The flower was not, in fact, larger than the flowers grown in previous iterations of the activity,[27] but the child appeared to believe otherwise. Here again, to the child's view there was a discrepancy between her current model of kind **FLOWER** and this new instance of *flower* that showed "different" size characteristics (a surprising event).

6.2 Specific category guidance

6.2.1 General note: Considering what children with autism may reasonably expect

Due to known difficulties with imitation in autism, including imitating actions on objects (e.g. as in Rogers et al., 2003), it also should *not* be automatically assumed that the child observing another person or a virtual character in a specific setting will lead to the child learning about the properties of objects, characters, etc. and forming expectations about their future properties and behaviour. Thus, the current scheme relies on the child's direct experience. However, this type of social learning certainly *may* occur, especially for more able children with autism or for typically developing children, in which case it may be appropriate to account for observation of others when determining how many times the child has been exposed to particular aspects (e.g. in determining reactions to novelty).

6.2.2 Novelty and the "rule of three"

As noted in Section 2.2.1, it is not yet clear how much experience is required in order to develop an expectation about some aspect of the environment. In order to make sure that novelty is classified consistently, some rule is required. The suggested guideline is the *"rule of three"*, which has the following considerations:

1. Child reactions to a novel aspect (both social and non-social) only have the potential to be "counted" as qualifying discrepancy-reaction pairs in the *first three times* the child experiences that aspect. Here, "experiences" means that the child could reasonably be expected to have seen, heard, or in some way attended to that aspect. If the child is otherwise occupied (i.e. speaking to the researcher about an unrelated topic, physically turned away from the screen, etc.) then the aspect should not be considered one of the three instances. In other words, it is assumed that the child requires three exposures in order to develop an expectation about a novel aspect as the first instance of a new kind. It is, of course, possible that the child will not perceptibly react to any of these three exposures (and thus, there are no analysable discrepancy-reaction pairs related to that aspect).

2. If the child continues to react to an aspect beyond the first three occurrences, the child is now assumed to have expectations about its behaviour and it is *no longer considered novel* for the purposes of annotation.

3. If an aspect is seen for the first time and violates physical laws, object relations, or other expectations that the child would be expected to have from outwith the current environment, it should still be treated as a potentially novel event on its first appearance, because it is still the first instance of a new kind.

4. If the child's reaction (or prediction, see Section 3.3.2) indicates that s/he *has already developed an expectation* about an aspect (and by extension, about that kind) in fewer than three exposures, then the "rule of three" ceases to apply. The child's reactions to that aspect should not be counted as discrepancy-reaction pairs, even if the child has experienced the aspect less than three times.

[27] It was possible to grow a range of flower sizes by shaking the "magic cloud" for longer or shorter periods of time, but a maximum size was set after which further cloud-shaking did not enlarge the flower.

6.2.3 Surprising events

No additional notes.

6.2.4 Non-events

The current data set suggests that recurrent [perceived] non-events may lead children to form new expectations, or rather, to update their previous expectations and expect that an item no longer has a particular property, etc. Should the object then become responsive again (often, in fact, the child imitates or discovers an effective touch-screen gesture) the child may then appear visibly surprised or amused at this violation of the new expectation. In the current data set, this was then counted as an additional discrepancy-reaction pair.

6.2.5 Child reactions

There is no maximum time limit that can elapse between a discrepancy and child reaction, but both should be within the same video segment (if there are multiple sessions and thus multiple pieces of video) unless there is extremely clear (likely verbal) evidence to suggest that in one session a child is referring to a specific discrepancy encountered in a previous session. There were no examples of this type in the current data set.

6.3 Impact of first-hand child knowledge on annotation

To what degree does first-hand knowledge of a participating child impact annotation? The answer is: to a surprisingly small degree in most cases. A good deal of child familiarity is obtainable from *spending time with the video data*, particularly by watching the child's entire sample through at least once before beginning any annotation. This can give a good sense of what the child is like, at least during ECHOES sessions, and his/her general style of interaction. Investing time with the videos prior to annotation also helps to alert the annotator about more specific issues, such as that a particular child frequently converses with the researcher but *never* looks away from the screen while doing so.

One place where first-hand knowledge can be important is in determining affect or engagement, especially for children who may appear very subdued compared to their peers, and/or have little facial expression. While some children may be obviously excited to any observer, exclaiming or moving around the room, another child's shifting in her chair may show clear excitement and interest *to a person who knows that child.* This type of information (e.g. "child may appear physically agitated, but is excited rather than frustrated or uncomfortable") can be conveyed to unfamiliar annotators through the use of child "profile" information completed by a teacher or familiar researcher.

An additional place where first-hand knowledge (or equivalent documentation) may be important is if the child makes non-standard use of verbal or sign language, in such a way that an "outside" annotator could not decipher it. For example, a child may have a customary phrase or sign that the classroom staff all understand as a request for help, but an outsider would not. This may be unique to the child, or part of a system used by the entire classroom or school (but not documented elsewhere).

7 Video annotation in ELAN

7.1 The annotation template

This section reports a procedure for applying the current taxonomies as an annotation scheme in ELAN. An example of a video in the process of being annotated is given in Figure 7.1. All annotation templates

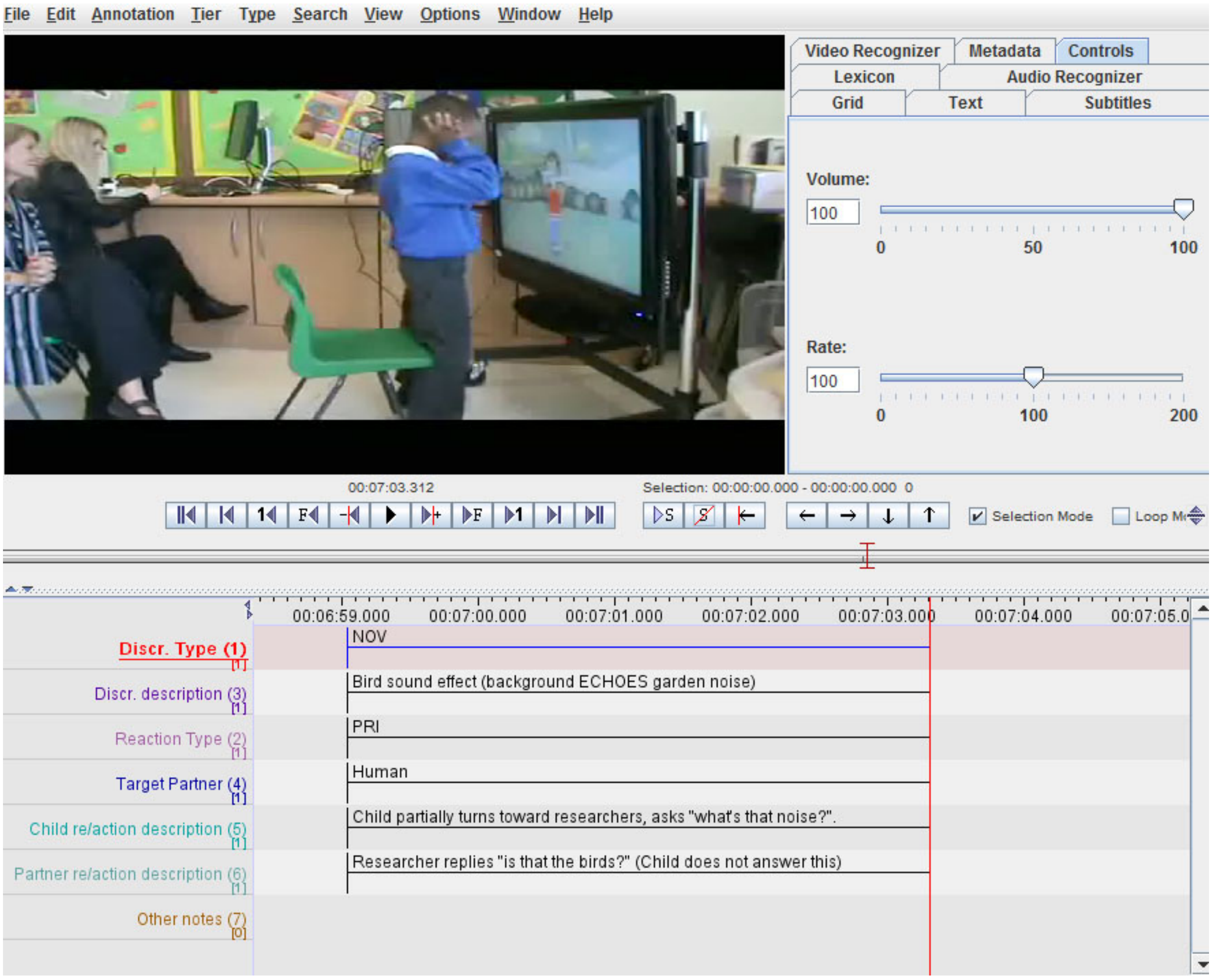


Figure 7: Example screen shot of video annotation in ELAN, showing annotations for child's primary initiation to a surprising event, and a response from the researcher.

were created by the author for the purpose of this analysis, and use independent tiers (rather than "parent" tiers with linked "child" tiers) for greatest flexibility.[28] Information is entered through a mixture of free-text entry and drop-down menus for applying predefined category labels (e.g. novelty, non-event, etc.). For examples, see Figure 8.[29]

The ELAN template for surprise-contingent child actions captured the following information about each event and subsequent child reaction, each on a separate tier:

1. **Discrepancy type:** A drop-down categorisation of the discrepant aspect as novel, surprising, or a non-event. This tier was left blank for predictions and child responses. There are separate options for these types where they pertain to secondary initiations.

2. **Discrepancy description:** A free-text description of the discrepancy, or a notation that the interaction concerned a preceding discrepancy (in the case of secondary initiations).
   NB: To assist in determining which events could be counted as novel, based on the "rule of three", sometimes the appearance of a new event or element would be noted *even if the child did not react to it,* but with the child reaction tiers (3-5) left blank. For example, the annotator might record "First instance of bubbles reward. No reaction."

[28]In ELAN, any tiers defined as child tiers are tied to the parent, annotations cannot be added unless an annotation already exists in that location on the parent tier. Parent tiers do not require child-tier annotations, and can stand alone.

[29]The latter are "controlled vocabularies" in the lexicon of the ELAN program.

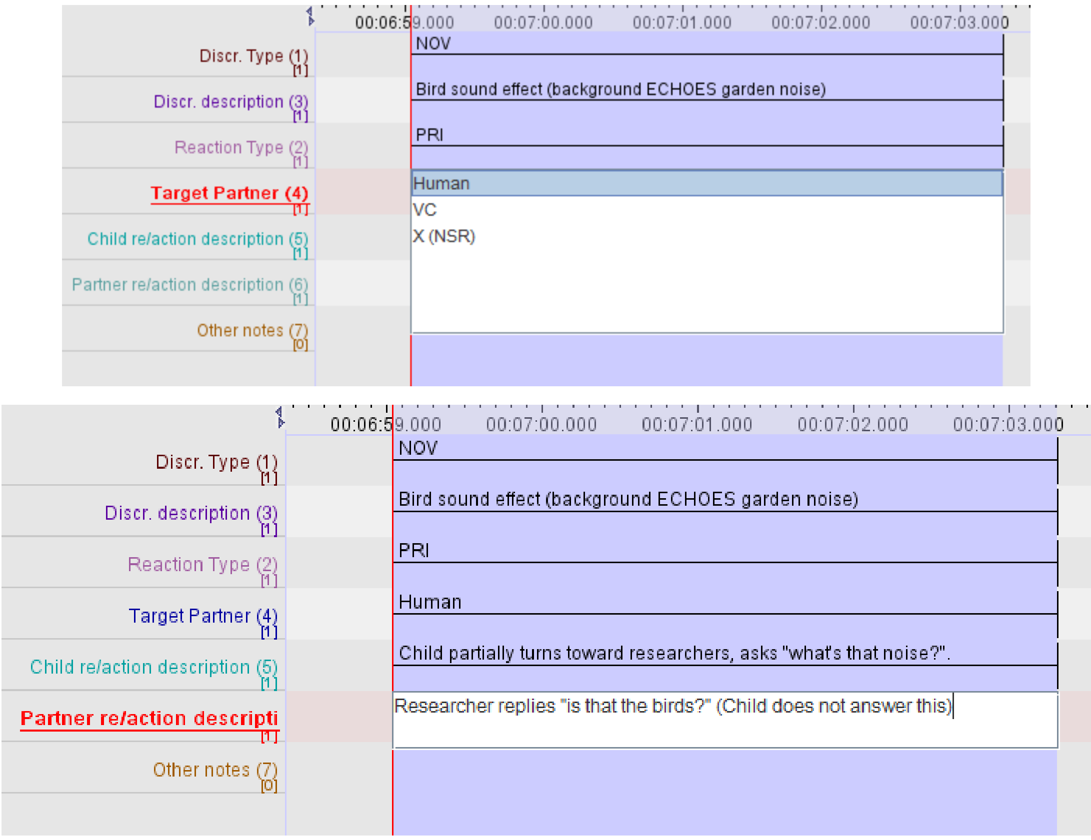


Figure 8: *Top:* Example of an annotation for a child's initiation to the researcher using a drop-down menu (controlled vocabulary). *Bottom:* Example of a free-text annotation about the researcher's verbal response to the child.

3. **Reaction type:** A drop-down categorisation of the child's reaction to the discrepant aspect. For simplicity, this tier also includes other child actions that are relevant (or potentially relevant) to the analysis, but that do not form part of discrepancy-reaction pairs: the labels for child predictions, and child responses to the partner.

4. **Target Partner:** The partner to which the child's initiation or response was directed (left blank for non-social reactions, predictions).

5. **Child re/action description:** A free-text description of the child's reaction to the discrepancy, or other relevant action (such as a prediction). Relevant child utterances were recorded in this tier, verbatim where at all possible.

6. **Partner re/action description:** A free-text description of the social partner's reaction to a child initiation (if any reaction). Relevant partner utterances were recorded in this tier, verbatim where at all possible. Also includes any other notes about relevant social partner actions that may be useful in determining reciprocal sequences, etc.

7. **Other notes**

Video segments were double-checked for consistency of taxonomy application, with the annotator watching them twice (with many stops and starts), looking for any missed discrepancy-reaction pairs or any erroneously annotated aspects or reactions. Finished annotation data was exported from ELAN as tab-delimited text and further analysed in a standard spreadsheet program.

7.2 Suggested annotation strategy

7.2.1 Three passes through each video

It is suggested to make three passes through each segment of video, in order to extract as much information as possible and improve the level of consistency within and across participants. The passes are as follows:

1. **Familiarisation:** Watch the full video segment at normal speed. If the segment for analysis does not start at the beginning of the video file, also watch the opening material.

2. **Annotation:** It is simply not possible to watch a video all the way through and label behaviours as it plays. In order to really understand what is occurring, the annotator will need to repeatedly stop and start, re-watching small segments multiple times in order to correctly identify them or make out the child's utterances.

   NB: In ELAN, it is also possible to slow down the video playback speed. Doing this pass at 90% speed can be very helpful.

3. **Additions and adjustments:** The goal for this pass is to check for possibly analysable segments missed in the second pass, and also to check and make sure than similar discrepancy-reaction pairs have been coded consistently throughout the video. It may be necessary to change codes to ensure consistency. This pass is *especially* important for each participant's first video segment.

It is strongly suggested to do all three passes for a given participant before moving on to the next participant. Alternately, an annotator might choose to familiarise her or himself with all participants' videos (pass 1) before beginning the annotation (passes 2 and 3).

After annotating several participants (perhaps four or so), it also may be wise to look again at the first participant's annotation, to check that similar types of aspects are being annotated consistently across children. To aid consistency, the annotator may wish to keep a running list indicating how certain types of aspects have been coded, and explaining the reasoning for coding particularly ambiguous instances in one way rather than another.

7.2.2 Keeping track of novelty

Section 6.2.2 describes the "rule of three" used to estimate whether or not a child has developed expectations about a novel aspect. In order for this rule to be most useful, the annotator should indicate in Tier 2 ("Discrepancy description" free-text field) whether a novel aspect is being seen/experienced for the first time, second time, or third time.

Keep in mind that a child may appear to attend to a novel aspect (e.g. observing a new character behaviour) but not make any observable reaction. However, the child may *still* react on a subsequent presentation. Thus, annotators are strongly advised to tag those initial presentations, by making a note in Tier 2, with the child reaction tiers (3-5) left blank. For example, the annotator might record "First instance of bubbles reward. No reaction."

7.2.3 Keeping track of secondary initiations

When annotating a secondary initiation, it is important to note the discrepant aspect to which it pertains, as the record for that initiation will be separated in time by some seconds (or potentially minutes) from the original aspect. It is strongly suggested to refer to a *description* of the aspect (e.g. "first time seeing fireworks") rather than an annotation number, as adding any further annotations during a subsequent pass through the video would then disrupt the numbering.

8 Spreadsheet-based analysis

In a discrepancy-reaction analysis (such as Alcorn et al., 2013), annotations were exported from ELAN as tab-delimited text (a .txt file) and opened in a standard spreadsheet program for ease of exploration, sorting, and summary. The spreadsheet-based analysis consisted of several tasks:

1. Revisiting the open-ended descriptions of the child's social reactions (i.e. those for which one of the social partners was identified as the target of the behaviour) to confirm that they constituted initiations, based on the definitions in the annotation scheme.[30] Those behaviours that did not qualify as initiations were re-tagged as non-social reactions (if applicable).

2. Qualifying initiations were tagged with their constituent child behaviours (e.g. pointing, gaze-shifting, commenting, etc.), for future analysis. Any number of tags could be applied.

3. Qualifying initiations were tagged for their function (imperative or social) based on the constituent child behaviours.

4. Any descriptions of non-social child reactions were tagged based on their apparent valence (positive, neutral/unclear, or negative).

5. Any sequences of initiations (see Section 3.4) were identified and their constituent lines of child and partner data copied to a separate "sequences" sheet for further analysis, namely checks for reciprocity of interaction.

6. Any explicit child predictions (see Section 3.3.2) were also copied to separate sheets for future analysis.

[30] Additional sources for identifying child initiations were the individual behaviours' presence in the initiation literature and other coding schemes. Of particular importance in this process were the list of goals in the SCERTS assessment process (SAP; Prizant et al., 2005), and the manual for scoring tasks in the Early Childhood Social Communication Scales (ECSC; Mundy et al., 2003). These are particularly helpful resources because they are very specific, and suggest clear "checks" on whether any given child action qualifies as an initiation, and which type it may be. Some, though not all, of the behaviours identified may constitute an initiation of joint attention. As the current project is interested in initiation more broadly, not solely in joint attention, this ambiguity is not currently a problem.

Aggregate spreadsheets (per participant and across all participants) reproduced key data for the purpose of determining behaviour counts in different categories and calculating descriptive statistics. As noted earlier, the results of an analysis of this type are reported in Alcorn et al. (2013).

Part III

Worked examples

9 Overview

Four examples of annotating children's discrepancy detection are reported at some length, in order to make clear how rich video material relates to the more abstract language of the taxonomy definitions and the steps in the decision tree diagrams. In conjunction with the notes about applying the taxonomy, these should be a reasonable tutorial in using the annotation scheme, or in beginning to adapt it to another context.

All of the following examples are drawn from the ECHOES data set, and their annotations explained step-by-step in accordance with how they were annotated and reported in Alcorn et al. (2013). The examples have been broken down and described step-by-step in terms of what the child, researcher and Andy (where he is relevant) are saying and doing. Each example ends at the point where the flow of activity moves away from the initial discrepant aspect, either because any interaction ceases entirely or because it is purposely re-directed to something else.

NB: All child names are pseudonyms

10 Example 1: Social reaction to novelty (New sound effect)

10.1 Context

Anthony is playing the "flower picking game" for the first time. With respect to the annotator locating discrepancies, this is an objectively new activity in the video sample with several objectively new aspects (a basket, a reward sound effect, several character utterances and behaviours). Thus, the annotator should be on "alert" that the child may likely react to novelty.

In this game, Andy asks for help putting flowers in the basket. He shows the child which flower he wants through one of three methods: gaze, pointing, or walking to the flower and touching it. Here, Andy has unsuccessfully indicated the first flower of the game through gaze, but now prompts the child again through gaze and pointing. Anthony touches the flower and it goes into the basket, causing a sound effect to play. It sounds like a short series of notes (~2 seconds) perhaps as played on an electronic keyboard. This sound effect is not used in any other activities; Anthony is hearing it for the first time in connection with ECHOES. Almost immediately after the sound effect, the researcher begins speaking to the child about one of the other objects in the activity, but Anthony is turned toward the screen and apparently not attending to this.[31]

10.2 Description of behaviours

1. Researcher: *Facing ECHOES control screen* "Let's–"

[31]There is no image included with this example because the discrepant aspect is auditory, and the child's orientation to the researcher at the control screen means that all that was visible to the camera was the back of the child's head.

Decision Tree Step	Step Outcome	Explanation
1. Does k appear to be an instance of a known kind **K**?	No, k is NOVEL (i.e. does not belong to a known kind).	The flower-picking game is being played for the first time and the child hears the sound effect (the "usemix") for the first time; no prior expectations can reasonably be assumed about either. Thus, the *flower-picking sound effect* is the first instance of new kind **MUSIC**.
Conclusion: Aspect was NOVEL. *Exit decision tree.*		

Table 1: Applying the discrepancy-type decision tree to the *flower-picking sound effect*

2. Anthony: *Turns to researcher* "Likes usemix!"
3. Researcher: "–leave the basket on the ground."
4. Anthony: *Continuing to face the researcher, he talks over her more loudly, waving his finger as though conducting an orchestra* "Likes usemix!" *He continues to watch the researcher.*
5. Researcher: *Still facing ECHOES control screen* "Yeah."
6. Anthony: *Still facing the researcher, he says more clearly* "Likes usemix."
7. Researcher: *Turns to child* "Usemix?"
8. Anthony: "Yeah." *He quietly imitates the game sound effect* "duh duh DUH duh DUH!"
9. Second researcher (behind camera): "He means the music!"
10. Researcher: *Smiles and laughs with Anthony, who is still looking at her.* "Ohhh. Yes, it is the music!"
11. Researcher: "[Andy's] going to show you again." *Tries to direct Anthony*'s *attention to screen.*

10.3 Applying the discrepancy type decision tree

Here, the aspect of the environment k to which Anthony appears to be attending and reacting is the *flower-picking sound effect* resulting from correctly choosing a flower. This could be considered as an instance of kind **SOUND EFFECT**, but he has clearly interpreted this as music, so it makes most sense to consider this as an instance of kind **MUSIC**. In this case, it is not crucial what precise label the kind is given.

We can determine whether this instance counts as discrepant, and what type of discrepancy it appears to be, by following the steps in the discrepancy type decision tree (Figure 2.3). An explanation of the relevant steps is documented in Table 1.

10.4 Applying the child reactions decision tree

In order to determine whether a discrepancy-reaction pair is present, we must begin by considering the *first* child behaviour that appears related to the discrepant aspect: the utterance "Likes usemix!" in line #2, above. This utterance is considered for inclusion in a discrepancy-reaction pair (and its type determined) by following the steps in the child reactions decision tree (Figure 3.5), which results in this action being recorded as a *primary initiation* (see also definitions in 3.1). Thus, there is a qualifying discrepancy-reaction pair (Novel event- primary initiation). A step-by-step explanation of this conclusion appears in Table 2. Any subsequent behaviours can then be considered afterwards individually, and may mean that additional discrepancy-reaction pairs are present.

Decision Tree Step	Step Outcome	Explanation
1. Is there a spontaneous, purposive reaction contingent on the occurrence (or non-occurrence) of k?	Yes. Spontaneous, purposive and contingent reaction to k.	Anthony has a visible and audible reaction immediately following the novel sound. The rest of the interaction underscores that he is definitely talking about the sound, and is concerned with making this understood (contingency). He was not prompted to attend to or react to this event (spontaneity).
2. Is the child's reaction clearly directed to a social partner?	Yes. It is a social reaction.	Anthony deliberately orients himself to the researcher and gazes at her as he makes his utterance. It appears clearly directed to her.
3. Is this the first reaction to that particular instance of k?	Yes. Child has made a PRIMARY INITIATION (PRI)	"Likes usemix!" is the first reaction, though the interaction continues thereafter. Those reactions would need to be considered separately.
Conclusion: A discrepancy reaction pair is present (Novel event-PRI). *Exit decision tree.*		

Table 2: Applying the child reaction decision tree to the child action in line #2 of example 1.

10.5 Dual taxonomy flowchart: Checking for further initiations and interactions sequences

10.5.1 What has been determined so far?

The steps undertaken so far have charted how an aspect of the environment has led the child to compare his mental model to the environment, and detect a "mismatch", or *discrepancy*. By using the discrepancy-type decision tree, an annotator can determine that this appears to be because the aspect of the environment was NOVEL, meaning that the child did not yet have knowledge or expectations about it. The steps in the child reactions decision tree tell us that Anthony made a primary initiation (PRI) to the researcher. There is thus at least one discrepancy-reaction pair present about this particular novel event.

The dual taxonomy flowchart (Figure 6) indicates that there are two "legal" states that could follow a PRI, which is to say two possible types of behaviours that continue to be about the original discrepancy. One is a partner response (PR) to the child's PRI, and one is an additional or *secondary* initiation from the child (SEC). From there, other options are possible depending on what happens next, including child response (CR), partner initiation (PI), and further SEC.

10.5.2 Subsequent child and partner actions

By consulting the child reactions decision tree several more times, it is possible to determine that Anthony is quite persistent in communicating about the "usemix" and appears aware that he has not been understood (or that the researcher is not really paying attention to him). He initiates multiple times, though is ultimately only understood because the second researcher speaks to the first researcher (whom the child has been addressing throughout the interaction), and the first researcher then continues the interaction with the child. This intervention is essentially external to the emerging interaction sequence between researcher and child, as the second researcher does not address the child directly and the child takes no notice of her, remaining completely focused on the first researcher.

Labelling the child and partner actions other than those in line #6 follows fairly straightforwardly according to the decision tree. Thus, only the child behaviour in line #6 is broken down in step-by-step in the same manner as the primary initiation, above, in order to illustrate why it is in fact a secondary initiation (SEC) instead of a child response (CR). This explanation is given in Table 3. A description of the communicative behaviours with their respective reaction-type labels is given below. Note that while partner responses (PR) are not listed in the decision tree, they follow the same criteria as child responses (see 3.3.1).

1. Researcher: *Facing ECHOES control screen* "Let's–" **(PI_1)**
2. Anthony: *Turns to researcher* "Likes usemix!" **(PRI)**
3. Researcher: "–leave the basket on the ground." **finishes (PI_1)**
4. Anthony: *Continuing to face the researcher, he talks over her more loudly, waving his finger as though conducting an orchestra* "Likes usemix!" *He continues to watch the researcher.* **(SEC_1)**
5. Researcher: *Still facing ECHOES control screen* "Yeah." **(PR_1)**
6. Anthony: *Still facing the researcher, he says more clearly* "Likes usemix." **(SEC_2)**
7. Researcher: *Turns to child* "Usemix?" **(PR_2)**
8. Anthony: "Yeah." *He quietly imitates the game sound effect* "duh duh DUH duh DUH!" **(CR)**
9. Second researcher (behind camera): "He means the music!" **(PI, second researcher to first researcher)**
10. Researcher: *Smiles and laughs with Anthony, who is still looking at her.* "Ohhh. Yes, it is the music!" **(PR_3)**
11. Researcher: "[Andy's] going to show you again." *Tries to direct Anthony's attention to screen.* **(PI_2)**

10.5.3 Checking for interaction sequences

The presence of multiple child initiations about the same discrepancy means that there is a *sequence* of interactions, as discussed in Section 3.4. The presence of at least one partner action directed to the child suggests that a reciprocal interaction sequence may also be present. Diagramming the actions sequentially in the form introduced in Section 3.4 allows an easy visual check for the presence of a reciprocal interaction sequence. Again, as noted in the discussion of sequences, there must be two or more consecutive "pairs" of contingent interactions between the child and the social partner (generally but not always one initiation and one response). At least one of these pairs must include an initiation from the child. There is no specific requirement for initiation from the partner.

Here, there are three pairs of child-researcher actions, preceded and followed by non-sequence actions. As noted in relation to Figure 4, arrows indicate contingency between actions. Note that there is no arrow between the PR in pair #1 and the SEC in pair #2: a secondary initiation (or any initiation) is definitionally *not* contingent on any preceding partner action (see Table 3 for a discussion of this specific SEC example). The series of contingency arrows continuing across pairs #2 and #3 show how those interactional "moves" continue to build on each other directly in terms of their form and/or content.

This example also illustrates how it is possible for only a subset of a longer interaction to qualify as fully reciprocal under the current definitions, even if it is preceded and/or succeeded by actions relevant to the initial discrepant aspect.

Decision Tree Step	**Step Outcome**	**Explanation**
1. Is there a spontaneous, purposive reaction contingent on the occurrence (or non-occurrence) of k?	Yes. Spontaneous, purposive and contingent reaction to k.	*See previous table.*
2. Is the child's reaction clearly directed to a social partner?	Yes. It is a social reaction.	Anthony remains oriented to the researcher and gazes at her as he makes his utterance. It appears clearly directed to her.
3. Is this the first reaction to that particular instance of k?	No, it is a subsequent reaction.	This is the third reaction, following PRI and SEC
4. Did the social partner respond to the child's previous reaction, make any other child-directed actions, or any actions related to OR make any other actions related to k?	Yes, and relevant partner action(s) attended to/ understood by the child.	Researcher says "yeah", which is still a partner response (PR) even though she does not appear to understand the child.
5. Was the child's social reaction contingent (i.e. directly dependent upon) the form/content of the partner's action(s)?	No, not contingent on form/content.	The child's action is related to the partner action in the sense that the child appears to speak again and continue the interaction because the researcher's utterance is not a "real answer" and suggests she is not paying attention to him or does not understand. The initiative for continuing the interaction clearly lies with the child. He does not "build on" the form or content of the researcher's response, and thus his action counts as a secondary initiation (SEC) rather than a (contingent) child response (CR).
Conclusion: An additional discrepancy reaction pair is present (Novel event-SEC). *Exit decision tree.*		

Table 3: Applying the child reaction decision tree to the child action in line #6 of example 1.

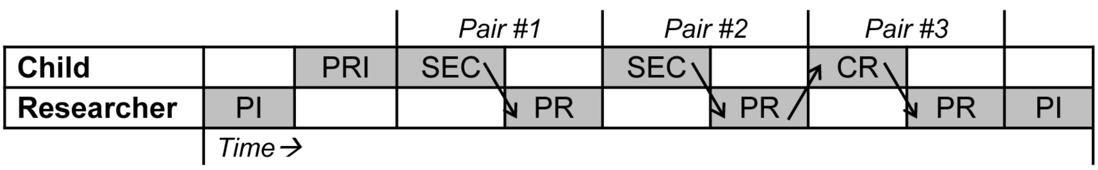


Figure 9: Diagram of reciprocal interactions in worked annotation example #1.

Figure 10: *Left:* Odell watches Andy make a sorting mistake. *Right:* Odell contact points to the blue box and tells Andy "Right here!".

10.6 Example 1 summary

The final items of interest annotated in this example would be as follows:

- 3 discrepancy-reaction pairs
  - 1 primary initiation (PRI) about NOVELTY, directed to the researcher
  - 2 secondary initiations (SEC) about NOVELTY, directed to the researcher
- 1 reciprocal interaction sequence with the researcher

11 Example 2: Social reaction to a surprising event (Andy's sorting mistake)

11.1 Context

Odell is playing the "ball sorting" game, in which he is supposed to help Andy sort bouncy balls (red, blue, and yellow) into the boxes of the same colours. This is his favourite ECHOES activity, and has been requested and played many times over multiple sessions. An annotator can reasonably assume that the child has formed expectations (though not necessarily correct ones!) about the objects, character behaviours, and goals of this activity. With respect to locating discrepancies, there are unlikely to be any objectively novel aspects in this much-repeated game. The annotator instead should be particularly watchful for surprising events and non-events.

In this iteration of the game, after Andy and Odell each sort several balls correctly, Andy picks up a blue ball and tries to put it in the yellow box (see Figure 10, Left). It rolls off the box top and bounces off the side of the screen. Andy turns to face forward (i.e. facing directly out at the child) and waits without saying anything.

11.2 Description of behaviours

1. Researcher: *begins* "Uh-oh, Andy–"
2. Odell: *Clearly focused on screen and not attending to researcher, he lunges forward and contact points at the blue box (i.e. touches it with extended index finger). He shouts* "Right here!" (See Figure 10, Right).
3. Researcher: *Continues* "–put it in the wrong box."

4. Odell: *Speaking over and still ignoring the researcher, he quickly switches hands and emphatically taps the blue box twice as he shouts* "That one!" *After a brief pause he shifts his attention to the other unsorted balls, and begins moving a red one.*

5. Researcher: "Can you show Andy how to do it, and put one in that box?" *Odell continues his own play with red ball at bottom of screen.*

While the researcher begins speaking about this event almost immediately, it seems fairly clear that the child is absorbed in his own interaction with Andy and is *not* attending to this utterance. This is an example of the issue mentioned in relation to the partner actions in the dual taxonomy flowchart (Section 4): there are many partner actions which may be relevant to a discrepant aspect, *but to which the child does not attend*. Thus, any child behaviours related to the same discrepancy should still be considered spontaneous and non-contingent on the partner behaviour, because the partner behaviour effectively does not exist as far as the child is concerned.

11.3 Applying the discrepancy type decision tree

There are several potentially relevant aspects in this example, and the first questions are which one(s) seems *most* relevant, and also to which extent the aspects should be grouped together. Which one is "driving" the child's reaction, so to speak? Given that Odell immediately appears to correct Andy and issue an instruction, it seems most likely that it is Andy's failed action to which he is reacting, rather than the behaviour of the ball (i.e. rolling off the top of the box). As noted above, Odell has played the sorting game many times and is thus well acquainted with the ball's properties (and has himself tried deliberately mis-sorting the balls to see what would happen). He has also seen Andy sort many balls before this one, using the same sequence of actions. Let us then treat the current instance as *Andy sorts ball into wrong box*, an instance of known kind or "category of things" **VC SORTING ACTIONS**. Prior to witnessing this mistake, the only instances of this kind with which the child would have been familiar were *Andy sorts ball into correct box*.

We can determine whether this instance counts as discrepant, and what type of discrepancy it appears to be, by following the steps in the discrepancy type decision tree (see Figure 2.3). An explanation of the relevant steps is documented in Table 4.

11.4 Applying the child reactions decision tree

In order to determine whether a discrepancy-reaction pair is present, we must begin by considering the *first* child behaviour: the utterance "Right here!" and contact point in line #2, above. As Odell does these simultaneously and the communicative forms reinforce one another, they are considered *as a unit,* rather than as two possible initiations. These actions are considered for inclusion in a discrepancy-reaction pair (and its type determined) by following the steps in the child reactions decision tree (Figure 3.5), which results in this action being recorded as a *primary initiation* (see definition in Section 3.1). Thus, there is a qualifying discrepancy-reaction pair (Surprising event- primary initiation). A step-by-step explanation of this conclusion appears in Table 2. *Any* subsequent behaviours can then be considered afterwards individually, should a discrepancy-reaction pair be present.

Decision Tree Step	**Step Outcome**	**Explanation**
1. Does k appear to be an instance of a known kind **K**?	Yes, k is an instance of kind **K**. It is FAMILIAR (i.e. non-novel).	Child has clearly documented prior experience with this game, the behaviour of the balls, and Andy's possible actions in this game, including **VC SORTING ACTIONS**.
2. Did k show the characteristics and behaviours that the child expected or predicted, based on knowledge about the structure of familiar kind **K**?	No, only a subset of the structure is present (i.e. some expectations violated). This is an instance of SURPRISE.	Andy performed a sequence of actions that was different than what was expected when he takes a turn at sorting, but still definitely constituted an instance of a **VC SORTING ACTION.**
3. In what way(s) did k differ from the known structure of **K**?	SURPRISING EVENT. Aspect k occurs or is present, but some of its characteristics or behaviours differ from **K** in a way that the child did not expect or predict...	*Andy sorts ball into wrong box* is clearly an event, rather than the lack of an event.
Conclusion: Aspect was a SURPRISING EVENT. *Exit decision tree.*		

Table 4: Applying the discrepancy-type decision tree to *Andy sorts ball in wrong box.*

Decision Tree Step	**Step Outcome**	**Explanation**
1. Is there a spontaneous, purposive reaction contingent on the occurrence (or non-occurrence) of k?	Yes. Spontaneous, purposive and contingent reaction to k.	Odell has a visible and audible reaction immediately following the surprising event. Even though the researcher begins to say something relevant to the event, Odell is not attending to this and can be considered to be acting spontaneously. He purposefully uses contact pointing to indicate a specific location, and makes two verbal utterances that based on the content of his utterances, seems clearly related surprising event (contingency).
2. Is the child's reaction clearly directed to a social partner?	Yes. It is a social reaction.	The utterance "Right here" appears to be a direction (or a correction!). Combined with the contact point to direct attention to a location in space *and* Odell's obvious disinterest in the researcher, he appears to be speaking to Andy.
3. Is this the first reaction to that particular instance of k?	Yes. Child has made a PRIMARY INITIATION (PRI)	The point plus utterance "Right here!" are the first reaction to the sorting mistake, though the interaction continues thereafter. Those reactions would need to be considered separately.
Conclusion: A discrepancy reaction pair is present (Surprising event-PRI). *Exit decision tree.*		

Table 5: Applying the child reaction decision tree to the child action in line #2 of example 2.

11.5 Dual taxonomy flowchart: Checking for further initiations and interactions sequences

11.5.1 What has been determined so far?

The steps undertaken so far have charted how an aspect of the environment has led the child to compare his mental model to the environment, and detect a "mismatch", or *discrepancy.* By using the discrepancy-type decision tree, an annotator can determine that this appears to be because the aspect of the environment was surprising and was an event (rather than a non-event), meaning that expectations were not met in some way. The steps in the child reactions decision tree tells us that Odell made a primary initiation (PRI) *to Andy.* There is thus at least one discrepancy-reaction pair present about this particular surprising event.

11.5.2 Subsequent child and partner actions

By consulting the child reactions decision tree again, we can see that Odell also makes a secondary initiation to Andy. Unfortunately, as noted in the context for this example, Andy mis-sorts the ball and then stands facing forward with no utterances or further actions. The researcher tries to initiate to the child about the discrepancy, but does not seem to successfully attract his attention. An annotated description of the communicative behaviours is as follows, inserting Andy's non-actions:

1. Researcher: *begins* "Uh-oh, Andy–" **(PI_1)**

   Andy: *Turns to face forward*

2. Odell: *Clearly focused on screen and not attending to researcher, he lunges forward and contact points at the blue box (i.e. touches it with extended index finger). He shouts* "Right here!" **(PRI)**

3. Researcher: *Continues* "–put it in the wrong box." **finishes (PI_1)**

   Andy: *Does nothing*

4. Odell: *Speaking over and still ignoring the researcher, he quickly switches hands and emphatically taps the blue box twice as he shouts* "That one!" *After a brief pause he shifts his attention to the other unsorted balls, and begins moving a red one.* **(SEC)**

   Andy: *Does nothing*

5. Researcher: "Can you show Andy how to do it, and put one in that box?" *Odell continues his own play with red ball at bottom of screen.* **(PI_2)**

   Andy: *Does nothing*

11.5.3 Checking for interaction sequences

The presence of multiple child initiations about the same discrepancy means that there is a *sequence* of interactions, as discussed in Section 3.4. However, the complete lack of interaction (or any action) from Andy, the relevant social partner, means that it is not possible for this to be a reciprocal interaction sequence. Only the child is trying to interact! There is also no possible sequence between the child and the researcher: even though she initiates twice, the child is not attending to her and thus does not interact. By diagramming the series of actions as in example 1, it is possible to see how the reciprocal interaction requirements are clearly unfulfilled both in the case of the child and Andy (11), and then again in the case of the researcher, to whom the child makes no response.

Child	PRI	SEC
Andy		
	Time→	

Figure 11: A non-reciprocal sequence (multiple initiations) in which there are no moves from the social partner, Andy.

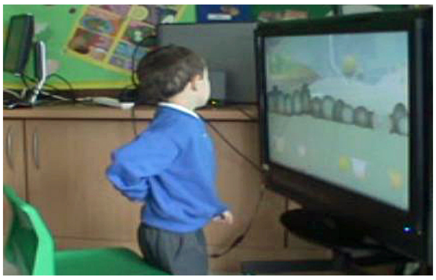


Figure 12: Ethan watches the screen for nearly 10 seconds, waiting for Andy to enter and demonstrate the activity. When Andy does not appear, he turns to the researcher for information (she is seated several feet behind him, beyond the right side of this image).

11.6 Example 2 summary

The final items of interest annotated in this example would be as follows:

- 2 discrepancy-reaction pairs
  - 1 primary initiation (PRI) about a SURPRISING EVENT, directed to Andy
  - 1 secondary initiation (SEC) about a SURPRISING EVENT, directed to Andy
- 1 non-reciprocal sequence

12 Example 3: Social reaction to a non-event (Andy does not enter)

12.1 Context

Ethan is standing in front of the screen, and the next activity (flower pot stacking) is about to begin. The screen fades to black, lights up, and the flower pots appear. Usually, Andy enters the scene approximately 5 seconds after the screen lights up. This time, approximately 10 seconds elapse without Andy entering. The ambient garden sounds continue as usual. During this time, Ethan pops the green "transition indicator" bubble that appears at the start of each new activity, then stands very still, watching the screen without touching it, as shown in Figure 12.

This is a discrepancy sequence that may be located by either of the methods suggested in Section 6.1: looking for an objective error, and monitoring a child's actions. Andy's failure to enter is definitely an objective error given the system's usual function, and Ethan's long period of watching and waiting is also unusual and worthy of investigation. What is he watching, or waiting for?

12.2 Description of behaviours

1. Ethan: *After about 10 seconds of waiting, he turns all the way around to look at the researcher, seated about 4 feet behind him. He continues looking until she responds (i.e. he is making a social reference).*

2. Researcher: *Gets up from her chair and takes several steps closer to Ethan. She leans down and says* "Wait 'til Andy will come back."

3. Ethan: *Orients himself back to the screen and waits.*

4. Andy: *Enters the screen, now about 15 seconds after the start of the activity.*

5. Researcher: "Look out" *she points to Andy, out of child's line of sight.*

6. Ethan: *Looks away from the screen and begins moving his chair.*

12.3 Applying the discrepancy type decision tree

As noted in Section 2.1.1, kinds are not only categories, but *sets* of features or constituent components that hold a certain relation to one another. In this case, the relevant kind is really the set of events that together constitute the beginning of a new ECHOES activity when they have a certain temporal relation to one another: The screen lights up, the ambient garden sounds begin again in the background, the activity's objects are present (or appear immediately), and then Andy enters on the left side of the screen after approximately 5 seconds. These components and their order stay the same regardless of which activity is loaded, and regardless of how many times an activity is played. Thus, we could call the current kind something like **ACTIVITY START SEQUENCE**. The relevant aspect appears to be Andy himself, or rather, his absence from this otherwise-familiar instance of **ACTIVITY START SEQUENCE**. His non-entrance (for though he eventually enters, at the time of the child's reaction he appears to be absent entirely) could be considered an instance of *Andy does not enter*.

We can determine whether this instance counts as discrepant, and what type of discrepancy it appears to be, by following the steps in the discrepancy type decision tree (see Figure 2.3). An explanation of the relevant steps is documented in Table 6.

12.4 Applying the child reactions decision tree

In order to determine whether a discrepancy-reaction pair is present, we must begin by considering the *first* child behaviour: Ethan's social referencing in line #1, above. *Social referencing* is a behaviour widely discussed in the developmental psychology literature and autism literature. It involves an individual (frequently an infant or small child) seeking information in a new or uncertain situation by looking to a social partner's face, for example to see if s/he looks frightened, encouraging, etc. Recall that, while worked examples 1 and 2 both discuss verbal (or verbal + non-verbal) child initiation behaviours, no verbalisation is required in order for a behaviour to count as a child initiation. This action is considered for inclusion in a discrepancy-reaction pair (and its type determined) by following the steps in the child reactions decision tree (Figure 3.5), which results in this action being recorded as a *primary initiation* (see definition in Section 3.1). Thus, there is a qualifying discrepancy-reaction pair (non-event- primary initiation). A step-by-step explanation of this conclusion appears in Table 7.

Decision Tree Step	Step Outcome	Explanation
1. Does k appear to be an instance of a known kind **K**?	Yes, k is an instance of kind **K**. It is FAMILIAR (i.e. non-novel).	As noted above, the activity start sequence is the same across all activities. The child already has extensive experience with ECHOES and is very familiar with the sequence and its timings.
2. Did k show the characteristics and behaviours that the child expected or predicted, based on knowledge about the structure of familiar kind **K**?	No, only a subset of the structure is present (i.e. some expectations violated). This is an instance of SURPRISE.	The components of **ACTIVITY START SEQUENCE** are all present and occur as usual, with the exception of Andy's entrance.
3. In what way(s) did k differ from the known structure of **K**?	[SURPRISING] NON-EVENT. Based on knowledge of kind **K**...k was absent when its presence was expected.	Andy was absent when his presence was expected. He makes a delayed entrance (after almost 3x as long as usual), but appears to be missing entirely at the time of the child reaction.
Conclusion: Aspect was a NON-EVENT. *Exit decision tree.*		

Table 6: Applying the discrepancy-type decision tree to *Andy does not enter.*

Decision Tree Step	Step Outcome	Explanation
1. Is there a spontaneous, purposive reaction contingent on the occurrence (or non-occurrence) of k?	Yes. Spontaneous, purposive and contingent reaction to k.	Ethan has a visible, un-prompted reaction following the non-event, which is to say he only reacts after enough time has elapsed that it is clear that Andy *should* be present, based on his knowledge of **ACTIVITY START SEQUENCE**.
2. Is the child's reaction clearly directed to a social partner?	Yes. It is a social reaction.	The fact that Ethan must turn his body almost 180° to see the researcher and holds his gaze until she begins to react makes the direction (and purposiveness) extremely clear.
3. Is this the first reaction to that particular instance of k?	Yes. Child has made a PRIMARY INITIATION (PRI)	Yes, no previous relevant child actions directed to a partner.
Conclusion: A discrepancy reaction pair is present (Non-event - PRI). *Exit decision tree.*		

Table 7: Applying the child reaction decision tree to the child action in line #1 of example 3.

12.5 Dual taxonomy flowchart: Checking for further initiations and interactions sequences

12.5.1 What has been determined so far?

The steps undertaken so far have charted how an aspect of the environment has led the child to compare his mental model to the environment, and detect a "mismatch", or *discrepancy*. By using the discrepancy-type decision tree, an annotator can determine that this appears to be because the aspect of the environment was surprising and was a non-event (rather than an event), meaning that expectations were not met in some way. The steps in the child reactions decision tree tell us that Ethan made a primary initiation to the researcher. There is thus at least one discrepancy-reaction pair present about this particular surprising event.

12.5.2 Subsequent child and partner actions

While the researcher makes two actions relevant to this discrepancy following the child's initiation (in lines #2 and #5), the child does not show any further communicative behaviours clearly related to the partner's action or the original discrepancy. Thus, there is no need to consult the dual taxonomy flowchart to check for child secondary initiations or responses to the partner. As a reciprocal sequence of interactions requires at least two child-partner exchanges (a minimum of four actions) there is also no need to check for sequences.

12.6 Example 3 summary

The final items of interest annotated in this example would be as follows:

- 1 discrepancy-reaction pair
  - Primary initiation (PRI) about a NON-EVENT, directed to the researcher
- No sequences

13 Example 4: Non-social reaction to novelty (Bubble reward)

13.1 Context

Hadi is playing the "ball sorting" for the first time. The annotator can assume that he has some expectations about the balls, as these have appeared in several previous activities which he has played over multiple ECHOES sessions. The other objects (boxes) are objectively new, as are Andy's activity-specific utterances and and movements (e.g. saying "let's put the balls into the boxes"). Thus, the annotator should be on "alert" that the child may likely react to novelty, though once the pattern of the activity becomes established, other discrepancies are of course possible.

In this game he is supposed to help Andy sort bouncy balls (red, blue, and yellow) into the boxes of the same colours (Example 2 also concerns this same activity). He has already completed the blue and yellow boxes, seeing a "fireworks" reward and an animated bumble bee reward, respectively. Hadi immediately moves on to completing the red box. He remains fully oriented to the screen throughout this sequence of behaviours.

Decision Tree Step	Step Outcome	Explanation
1. Does k appear to be an instance of a known kind **K**?	No, k is NOVEL (i.e. does not belong to a known kind).	The ball sorting game is being played for the first time, and the rewards seen for the first time. While the child is already familiar with bubbles as an object, no prior expectations can reasonably be assumed about the type of reward that would follow completion of the red box. Thus, *bubbles appear* is the first instance of new kind **RED BOX REWARD**.
Conclusion: Aspect was NOVEL. *Exit decision tree.*		

Table 8: Applying the discrepancy-type decision tree to *bubbles appear*.

13.2 Description of behaviours

1. Andy: *Gives thumbs up and says* "good job" *(in relation to completing the yellow box and seeing the bee reward).*
2. Hadi: *Sorts third and final red ball into the box. It turns into three white bubbles.*
3. Hadi: *Pops the bubbles, quietly exclaiming* "Oh! Oh!" *as he does the second and third.*
4. Andy: *Looks from side to side around the now-empty garden. He says* "All done" *and then stands and waits for several seconds before turning and walking off the right side of the screen. Hadi continues to quietly watch Andy until the activity is over and a new one loads.*

13.3 Applying the discrepancy type decision tree

Here, the aspect of the environment k to which Hadi appears to be attending and reacting is clearly the appearance of the bubbles (i.e. *bubbles appear).* This can be considered an instance of the kind **RED BOX REWARD**. We can determine whether the instance *bubbles appear* counts as discrepant, and what type of discrepancy it seems to be, by following the steps in the discrepancy type decision tree (see Figure 2.3). An explanation of the relevant steps is documented in Table 8.

While the child is already familiar with bubbles and their properties from other ECHOES activities, the *relation* between the appearance of the bubbles and the completion of the sorting task is new, and no prior expectations can reasonably be assumed about the type of reward that would follow completion of that box. As discussed in Section 2.1.1 and also as part of example 3, kinds are not only conceptual categories but relations between features and/or components. Here, the kind **RED BOX REWARD** is composed of both temporal and causal relationships between the balls, the box, and the appearance of the bubbles. It is this relationship that is not yet part of the child's mental model of the environment, even though the individual components are already familiar.

13.4 Applying the child reactions decision tree

In order to determine whether a discrepancy-reaction pair is present, we must consider the only child behaviour in this sequence that appears related to the discrepant aspect: the exclamation in line #3, above. This utterance is considered for inclusion in a discrepancy-reaction pair (and its type determined) by following the steps in the child reactions decision tree (Figure 3.5), which results in this action being recorded as a *non-social reaction* (see also definitions in 3.2). Thus, there is a qualifying discrepancy-reaction pair (Novel event- primary initiation). A step-by-step explanation of this conclusion appears in Table 9.

Decision Tree Step	Step Outcome	Explanation
1. Is there a spontaneous, purposive reaction contingent on the occurrence (or non-occurrence) of k?	Yes. Spontaneous, purposive and contingent reaction to k.	Hadi's verbal reaction seems connected to the bubbles' appearance, especially as synchronises his "Oh! oh!" with popping the second and third bubbles. This is clearly unprompted because neither possible social partner attempts to interact with him during this sequence.
2. Is the child's reaction clearly directed to a social partner?	No. It is a NON-SOCIAL REACTION (NSR)	Hadi remains completely oriented to the screen throughout this sequence and for some time after. He neither acknowledges the researcher's presence, nor attempts to interact with Andy.
Conclusion: A discrepancy reaction pair is present (Novelty-NSR). *Exit decision tree.*		

Table 9: Applying the child reaction decision tree to the child action in line #3 of example 4.

13.5 Dual taxonomy flowchart: Checking for further initiations and interactions sequences

13.5.1 What has been determined so far?

The steps undertaken so far have charted how an aspect of the environment has led the child to compare his mental model to the environment, and detect a "mismatch", or *discrepancy*. By using the discrepancy-type decision tree, an annotator can determine that this appears to be because the aspect of the environment was NOVEL, meaning that the child did not yet have knowledge or expectations about it. The steps in the child reactions decision tree tell us that Hadi made a non-social reaction (NSR), meaning that it was not directed to a social partner. There is thus at least one discrepancy-reaction pair present about this particular novel event.

13.5.2 Subsequent child and partner actions

The dual taxonomy flowchart indicates that it is possible to have additional discrepancy-reaction pairs if the first child reaction is an NSR: the child may then initiate to a partner about the discrepancy (constituting a PRI). However, in this case there are no further child actions of any kind before the start of the next activity. Thus, there is no need to continue through the flowchart or check for sequences.

13.6 Example 4 summary

The final items of interest annotated in this example would be as follows:

- 1 discrepancy-reaction pair
  - Non-social reaction (NSR) about NOVELTY.
- No sequences

14 Summary

The four worked examples in this section cover the three types of discrepancy (novelty, surprising events, and non-events) and a range of social and non-social child reactions. They also illustrate several tricky annotation issues, such as determining whether a subsequent child reaction after a partner response should count as a secondary annotation or a child response (example 1), or identifying kinds that are novel because of the *relations* between their component objects/actions rather than the components themselves (example 4). These examples are certainly not exhaustive, but are meant to give context for the taxonomies and to show how the decision trees and dual taxonomy flowchart can be useful tools for “breaking down” how a child is interacting with the environment and social partners.

The worked examples are presented at length in the hope of assisting, rather than overwhelming, the reader/annotator. The first sessions of annotating a data set are by far the slowest, with the most explicit decision-making steps and reference to these tools. Later, when there are many duplicate instances, many of the decisions about how to “count” things will have already been made and the process speeds up enormously. For example, first encountering the bubbles reward during the sorting task will be an instance of novelty for *all* children [who react to it]. The annotation challenge then becomes maintaining consistency over time. Spreadsheet analysis (see Section 8), in which it is possible to sort the finished annotations and examine all instances of a certain discrepancy type or child reaction, is invaluable for checking consistency across participants (or even videos within participants). Some adjustments at the spreadsheet stage are inevitable.

As a final recommendation, it is strongly, strongly suggested to keep annotation notes, particularly about how complex examples were categorised. In some cases, more than one relevant rule may seem to apply, and it is up to the annotator to decide which is the most important or most relevant, and then apply the spirit of that decision consistently throughout the data set.

References

(2012). *ELAN Linguistics Annotator, version 4.4.0.* Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands.

Alcorn, A., Pain, H., and Good, J. (2013). Discrepancies in a virtual learning environment: something worth communicating about for young children with ASC? In *Proceedings of the 12th International Conference on Interaction Design and Children*, pages 56–65. ACM.

Alcorn, A., Pain, H., Rajendran, G., Smith, T., Lemon, O., Porayska-Pomsta, K., Foster, M. E., Avramides, K., Frauenberger, C., and Bernardini, S. (2011). Social communication between virtual characters and children with autism. In Biswas, G., Bull, S., Kay, J., and Mitrovic, A., editors, *Artificial Intelligence in Education*, volume Lecture No, pages 7–14, Berlin. Springer Verlag.

Bernardini, S. and Porayska-Pomsta, K. (2013). Planning-Based Social Partners for Children with Autism. In *Proceedings of the Twenty-Third International Conference on Automated Planning and Scheduling Planning-Based*, pages 362–370.

Boden, M. (1995). Creativity and Unpredictability. *Stanford Humanities Review*, 4(2):1–18.

Boden, M. (2010). *Creativity and Art: Three Roads to Surprise.* Oxford University Press, Oxford, UK.

Boden, M. A. (1990). *The Creative Mind: Myths and Mechanisms.* Abacus, London, 1992 edition.

Foster, M., Avramides, K., Bernardini, S., Chen, J., Frauenberger, C., Lemon, O., and Porayska-Pomsta, K. (2010). Supporting Children's Social Communication Skills through Interactive Narratives with Virtual Characters. *Proceedings of the international conference on Multimedia*, (Figure 2):1111–1114.

Goswami, U. (1998). *Cognition in Children.* Psychology Press/Erlbaum (Uk) Taylor & Francis.

Hausman, C. R. (1975). *A Discourse on Novelty and Creation.* Martinus Nijhoff, The Hague, Netherlands.

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., and Seibert, J. (2003). A Manual for the Abridged Early Social Communication Scales (ESCS).

Porayska-Pomsta, K., Frauenberger, C., Pain, H., Rajendran, G., Smith, T., Menzies, R., Foster, M. E., Alcorn, a., Wass, S., Bernadini, S., Avramides, K., Keay-Bright, W., Chen, J., Waller, a., Guldberg, K., Good, J., and Lemon, O. (2011). Developing technology for autism: an interdisciplinary approach. *Personal and Ubiquitous Computing*, 16(2):117–127.

Prizant, B., Wetherby, A., Rubin, E., Laurent, A., and Rydell, P. (2005). *The SCERTS [TM] Model: A Comprehensive Educational Approach for Children with Autism Spectrum Disorders.* Brookes Publishing Company, Baltimore, MD.

Rogers, S. J., Hepburn, S. L., Stackhouse, T., and Wehner, E. (2003). Imitation performance in toddlers with autism and those with other developmental disorders. *Journal of Child Psychology and Psychiatry*, 44(5):763–781.

White, P. J., O'Reilly, M., Streusand, W., Levine, A., Sigafoos, J., Lancioni, G., Fragale, C., Pierce, N., and Aguilar, J. (2011). Best practices for teaching joint attention: A systematic review of the intervention literature. *Research in Autism Spectrum Disorders*, 5(4):1283–1295.

Appendix D

Ethics and information forms for ECHOES

Study 1, reported in Chapter 5, applies a new analysis to data collected as a part of the ECHOES project summative evaluation. This use is permitted by the terms of the original project ethics application and consent paperwork, excerpts or copies of which are provided in this appendix.

D.1 Consent to data access and use

Parents of participants in the evaluation (all sites) consented to study participation, including video collection and standardised measures, with data "access [to] be limited to the people involved in this research" (see parent information, in this appendix). This includes the author, who helped to collect evaluation study data in her role as a project researcher, and who has been an ongoing contributor to ECHOES analyses and publications during the PhD period. Due to the broad, interdisciplinary, and exploratory nature of the ECHOES project, it was anticipated that there would be multiple analyses by different project researchers over time, for different audiences. Thus, the ethics application and consent paperwork refer to high-level research goals rather than specific uses of the collected data. The current research is congruent with the goals as described in the parent information sheet, namely development of "game-like learning software that children can use to explore and improve their social interaction and communication skills". In the current analyses, data has been stored and participant identities protected in accordance with the terms of the ECHOES ethics application and parent information and consent. Images of children have been included in the thesis and related publications in line with parents' additional consent (where provided) regarding video and image use.

D.2 Documents

D.2.1 ECHOES ethics application

As noted in the main thesis text, the ECHOES project, including work by associated research students, received ethical approval from the Psychology, Philosophy, and Linguistics (PPLS) ethics committee at the University of Edinburgh (2008; Reference: 160-0708). All other ECHOES partner institutions agreed to accept ethical approval from the University of Edinburgh, in lieu of application to their own local ethics committees. The sections of the original ECHOES ethics application most pertinent to the current research use are reproduced verbatim as follows:

SECTION 3: RESEARCH DESIGN AND PROTOCOL[1]

3.8 What information will you give to participants prior to their consent?

Once prospective participants who meet the study criteria are identified, they will be invited to participate by our sending their parents (or relevant carer) information about the project. Parents will be provided with a written information sheet about the study, a consent form and a pre-stamped and addressed envelope. The information sheet will contain an invitation to take part in the study and will describe in lay language (1) the purpose of the study, (2) the structure and duration of the sessions, (3) a statement describing the extent to which confidentiality of records identifying the participant will be maintained, (4) the voluntary nature of participation, (5) the participant's right to withdraw from the study at any time without penalty and to have his/her information destroyed, and (6) consent to use data and video for teaching, research and dissemination purposes.

SECTION 5: DATA PROTECTION

5.1 Will any part of the research involve audio or film recording of individuals?

Data will include performance evaluation using various metrics, video and audio recordings of interactions, observational data, interviews, questionnaires and verbal protocols.

[1] Sans-serif text indicates verbatim text from ECHOES documentation.

General information such as name, age, gender, and relevant test scores and other data from school records will be collected from each participant. Participants and parents will be asked for consent to use data and video for teaching, research and dissemination purposes.

5.4 How will confidentiality of participants be protected?

Participants will be identified exclusively by an ID code throughout documentation (except where their names naturally occurred in the speech stream).

5.5 Who will be entitled to have access to the raw data? What steps have been taken to ensure that only entitled persons will have access to the data?

Data access will be limited to ECHOES II researchers (and associated research students, with permission: see study document).[2] The identities of participants will be protected unless specific permission is given for dissemination purposes. There will be limited dissemination of recorded data, as agreed with participants, for teaching, research and dissemination purposes. If data is to be released outside ECHOES II, a second approval process will be carried out for all participants involved. This will be monitored and subject to the ECHOES II ethics committee (see study document).

5.6 How and where will the data be stored, in what format, and for how long?

All research data will be analysed and stored, in accordance with data protection guidelines, at partner institutions responsible for collecting them. Physical data, such as videos, questionnaires and transcriptions, will be stored in a locked cupboard at each institution responsible for collecting them. Privacy and security of electronic data will be achieved through the extensive security and data protection software that each institution has. Anonymous logs of data collected will be publicly available through the ESRC and the project website. Data will be used for agreed purposes only. We will hold the data only for the time of research, and as long as the scientific community can benefit from these data.

[2] This document described the different types of studies intended to be undertaken during the ECHOES project, e.g. design workshops, evaluation studies.

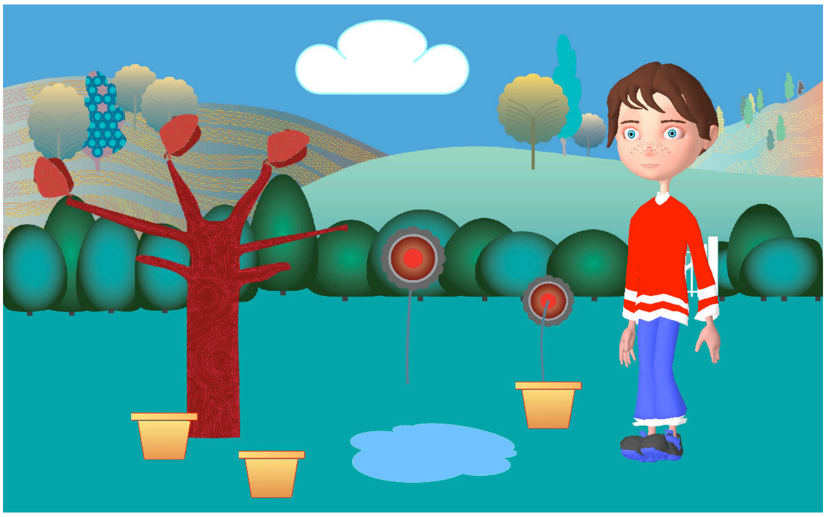
D.2.2 Parent information and consent

The parent information and consent forms from the ECHOES summative evaluation are included at the end of this appendix. Regarding data use, please see especially the first paragraph on page 1 of the parent information, stating the high-level purpose of the data collection, and the section “How will information be stored?” on page 2 of the same document.

ECHOES

Dear Parent or Guardian,

We would like to introduce you to the ECHOES research project. We are developing game-like learning software that children can use to explore and improve their social interaction and communication skills, using a large multi-touch screen. The children explore a 'magic garden' and play with a child-like character called Andy.



We are carrying out research in your child's school and we would like your child to participate in this. This letter tells you about the project and the study we plan to run. If, after reading this you still have any questions, or want more detailed information, please feel free to ask us. You will find contact details below.

When and where will it happen?

Children will work with ECHOES in school, during class times, for eight 20-minute sessions over May and June (ideally starting in the next few weeks).

At the start of each session, the researcher will explain to the child what will happen, and will give them a chance to ask questions. They will tell the child that they can take a break or stop at any time without having to give a reason. The child will then play with ECHOES for 20 minutes, at most, then return to their class.

As the child is using a touch screen, they do not need to have well developed language skills, nor be able to read instructions – all instructions will be spoken.

What else does taking part in the study involve?

At the start and end of the study we will take 4 short videos (about 5 to 10 minutes) of your child: in the classroom, working one-on-one or in a group with a researcher or teacher and during playtime. We will also video record your child when working with ECHOES. These videos will be the main information collected for this research (it is not possible for us to replace video collection with extensive note-taking).

We will also need to collect background information about your child. We will use the British Picture Vocabulary Scale (BPVS), a widely used test of a child's verbal understanding of language. This takes 5 to 8 minutes to administer and will indicate your child's language understanding (your child's school may already use this).

If your child already has a diagnosis of ASD, we will also ask you to complete the Social Communication Questionnaire (SCQ) to give us a clearer picture of them. This is a short questionnaire for parents/guardians made up of 40 questions about your child's behaviour. It takes less than 10 minutes to complete. If SCQ forms are missing or incomplete, an ECHOES researcher may phone you to complete the questionnaire or to clarify some of your responses.

How will information be stored?

The videos and information will be stored safely and kept separate from information about your child's identity. Access will be limited to the people involved in this research.

We would like to include video clips, as examples, in research presentations or for teaching. In these, no reference to the child's identity will be made. If a photograph or video-clip is used, their name will be changed. If you prefer that we do not use the videos of your child for these additional purposes, you can indicate this on the consent form.

Would you like to participate?

If you and your child agree to take part in the ECHOES research project, please complete and sign the attached consent form and return it to school. Also please read the child assent form aloud to your child and help them to sign it: this is to let you child know that they can stop the study at any time, without having to give any reason. We will also remind the child at the start of each session that they can take a break or stop at any time.

Note that you may withdraw your child from the study at any time and for any reason simply by contacting your child's teacher or the researchers. Researcher contact information is listed at the end of this form.

Further information and queries:

Systems like ECHOES are not yet available to buy, though one day they may be. These studies, and your child's help with them, will provide ideas about how teaching and communication systems may be improved in the future.

Thank you for taking the time to read this. If you would like to know more about this research and/or if you have questions, please feel free to contact Dr Helen Pain from the School of Informatics, (helen@inf.ed.ac.uk) or on 07974 971 475.

You can also find out more at our website, http://echoes2.org/, where there is also a short video.

Thank you very much.
The Research Team of the ECHOES Project

ECHOES Project: Consent Form for Parents

	Please circle
Have you read the covering letter?	YES / NO
Have you received enough information about the project?	YES / NO
Do you understand that participation is completely voluntary and your child can withdraw at any time without having to give a reason?	YES / NO
Do you consent for your child to take part in this project, including video recording and background data collection through the British Picture Vocabulary Scale (BPVS)?	YES / NO
Do you consent to completing the Social Communication Questionnaire (SCQ) [1] and for ECHOES researchers to contact you by telephone if the SCQ is incomplete or has not been returned?	YES / NO

With reference to further use of photographic or video data (see the covering letter for details), please circle yes or no in response to the following:

> I AGREE that my child's likeness (e.g. on videotape) can be used as examples in research papers and presentations. YES / NO

Please date and sign this page below to indicate that you understand and accept the conditions of this study. Thank you.

Name of child participating in study: ___________________________________

Child's D.O.B: **…../…../……**

Your relationship to the child: ___________________________________

Your name (please print): ___________________________________

Contact telephone number: ___________________________________

Best time to reach you? ___________________________________

Signature: ___________________________________

Date: _____ / _____ /______

[1] Only relevant if your child has a prior diagnosis of ASD

Appendix E

List and description of aspect groups from Study 1 (ECHOES)

E.1 List of aspect groups and constituent aspects

There are 112 unique ECHOES *aspects* which were part of a discrepant situation in some way, or, in other words, could be considered to have "caused" a discrepancy-reaction pair. These are organised into 19 mutually exclusive *aspect groups*, a higher level of description that grouped apparently related aspects. Aspect groups are then in turn organised by their discrepancy subtype (novelty or surprise). Aspect groups, and aspects within groups, are all listed and numbered alphabetically for ease of reference.

Numbers given in parentheses refer to the number of unique discrepant instances with that aspect label.[1] For example, "Andy demonstrates stacking a flowerpot (3)" means that on three occasions in the dataset children reacted to this event as being novel. In total, there are 315 unique discrepancies in the dataset. Not that the number of total child reactions is higher, due to the presence of secondary initiations about some aspects.

E.1.1 Novel aspect groups

1. Andy action resulted in ball-sorting activity reward (total = 5) 1.1. Andy action resulted in ball-sorting activity reward: Bees (2)

[1]This is not the same as the number of DR pairs, as there may be multiple DR pairs about the same instance, as in the worked annotation example at the end of Chapter 4

1.2. Andy action resulted in ball-sorting activity reward: Bubbles (1)

1.3. Andy action resulted in ball-sorting activity reward: Fireworks (2)

2. Andy demonstrated novel action, object property, or relationship (total = 17) 2.1. Andy demonstrated sorting a ball into a box (2)

2.2. Andy demonstrated throwing a ball through the cloud (4)

2.3. Andy demonstrated shaking cloud to grow flower (7)

2.4. Andy demonstrated picking a flower (1)

2.5. Andy demonstrated stacking a flowerpot (3)

3. Andy made a novel utterance (total = 6) 3.1. Andy said “That’s fun” (1)

3.2. Andy said “Woah” (4)

3.3. Andy said “Wow!” (1)

4. Child action resulted in ball-sorting activity reward (total = 13) 4.1. Child action resulted in ball-sorting activity reward: Bees (4)

4.2. Child action resulted in ball-sorting activity reward: Bubbles (4)

4.3. Child action resulted in ball-sorting activity reward: Fireworks (5)

5. Child formed expectations about Andy (total = 20) 5.1. Child discovered how to tickle Andy (9)

5.2. Child met Andy (6)

5.3. Child saw Andy exit screen at end of activity or session (5)

6. Child discovered novel object action, property, or relationship (total = 36)

6.1. Child discovered how to bounce a ball (1)

6.2. Child discovered how to make the magic leaves from the tree fly around (2)

6.3. Child discovered how to pick up and throw a ball (1)

6.4. Child discovered how to sort a ball (3)

6.5. Child discovered moving and popping bubbles (1)

6.6. Child discovered shaking cloud to grow flower (8)

6.7. Child discovered that balls change colour when thrown through cloud (5)

6.8. Child discovered that flowerpots can be unstacked (1)

6.9. Child discovered that flowerpots change colour when stacked (1)

6.10. Child discovered that flowers can be wiggled on their stems, causing them to sway (2)

6.11. Child discovered that flowers have no gravity and can hover in the air (4)

6.12. Child discovered that it is (also) possible to move flowerpots by flicking them (2)

6.13. Child discovered that it is possible to flick balls out top of screen (2)

6.14. Child discovered turning a flower into a ball (1)

6.15. Child discovered turning a flower into a bubble (2)

7. Child was presented with a new scene (total = 13)

7.1. Child encountered the tree (3)[2]

7.2. Child perceived that there was a "new garden" (4)[3]

7.3. Child presented with configuration for ball-sorting (2)

7.4. Child presented with configuration for throwing balls through the cloud (1)

7.5. Child presented with configuration for growing flowers in pots (1)

7.6. Child presented with configuration for turning flowers into balls (1)

7.7. Child presented with the garden backdrop for the first time (1)

8. Novel sound effect (total = 12)

8.1. Novel sound effect: ball bouncing (2)

8.2. Novel sound effect: birdsong, as part of garden background sounds (2)

8.3. Novel sound effect: fanfare when correct flower picked (4)

8.4. Novel sound effect: static from system speakers (1)

8.5. Novel sound effect: thunder when cloud shaken repeatedly (3)

E.1.2 Surprising aspect groups

9. Andy interfered with the child's turn in an activity (total = 8)

9.1. Andy interfered with pot-stacking (3)

9.2. Andy stole ball from child (4)

9.3. Andy stole flower from child (1)

10. Andy was unresponsive to the child's communication attempt (total = 9)

10.1. Andy is unresponsive to child's greetings (6)

10.2. Andy did not respond to face/body being poked (2)

10.3. Andy did not respond when child introduced himself (1)

11. Andy violated a pattern/rule regarding an activity's rules/goals (total = 13)

11.1. Andy picked up two balls at once (1)

11.2. Andy sorted a ball into the wrong box (5)

11.3. Andy stacked two pots at once (1)

11.4. Andy stole an already-sorted ball out of the box (3)

11.5. Andy threw a ball at the cloud and missed (1)

---

[2]The tree is treated as a part of the scenery, rather than an object, because these instances are children reacting to its appearance alone (e.g. exclaiming "a tree!" when the scene lit up) rather than reacting to any interaction with it. In fact, the tree trunk itself was completely static.

[3]One of the two children who was convinced that the "old garden" was gone also talked repeatedly about the "new garden".

11.6. Andy tried to sort two balls at once (2)

12. Andy violated a pattern/rule regarding his own usual behaviour (total = 29)

12.1. Andy did not respond to tickling or tickle action was incorrect (4)

12.2. Andy exited unexpectedly or inappropriately (3)

12.3. Andy was stuck partly on screen, right (1)

12.4. Andy was stuck with his eyes closed (1)

12.5. Andy was inactive for a long period (8)

12.6. Andy was perceived absent at the start of the activity (7)

12.7. Andy would not accept an offered object (3)

12.8. Perceived missing behaviour: Andy not using ”stop” hand signal (1)

13. Child could not perform an expected object action: non-touchscreen cause (total = 19)

13.1. Could not perform impossible but “expected” action: push ball completely out of screen side (1)

13.2. Could not perform impossible but “expected” action: put flying leaves back on the tree (2)

13.3. Could not perform impossible but “expected” action: open gate in garden backdrop (1)

13.4. Child touched ball to magic cloud; it should change colour but did not (1)

13.5. Child tried to perform disabled action: turn flower into a ball (9)

13.6. Child tried to perform disabled action: pop red transition bubble (4)

13.7. Child misunderstood cause-effect relationship, tried to use a flower to make rain (1)

14. Child could not perform an expected object action due to touchscreen or gesture issue (total = 60)

14.1. Child could not move or pick up object at all: balls (17)

14.2. Child could not move or pick up object at all: flowerpot (14)

14.3. Child picked up object but it kept dropping or ricocheting repeatedly: ball (10)

14.4. Child picked up object but it kept dropping repeatedly: flowerpot (3)

14.5. Child’s name on start screen did not respond to touch or child action was incorrect: (1)

14.6. Object did not respond to touch or child action was incorrect: bubble (2)

14.7. Object did not respond to touch or child action was incorrect: cloud (7)

14.8. Object did not respond to touch or child action was incorrect: flower (6)

15. Child perceived that an aspect was missing (total = 11)

15.1. Perceived missing object: balls (1)

15.2. Perceived missing object: bubbles (1)

15.3. Perceived missing object: cloud (5)

15.4. Perceived missing object: flowers (1)

15.5. Perceived missing object: unclear, but child definitely convinced something was missing (1)

15.6. The "old garden" was perceived to be missing (2)This sounds very strange given that the garden background was the same throughout, but two children were at some point asked about where the "old garden" had gone.

16. Mismatch between partner utterance and environment state (total = 8)

16.1. Andy said activity was "all done" when it clearly was not (1)

16.2. Andy said the ball changed colour but pointed to one that had not (1)

16.3. Andy said he would do a particular action, but did nothing (2)

16.4. Researcher said Andy would enter or "come back" but he did not (2)

16.5. Researcher said Andy would take his turn or do a certain action, but he did not (2)

17. Unexpected or altered property of a familiar object (total = 17)

17.1. Ball makes huge bounces and flies out of top of screen (1)

17.2. Object somehow did not behave as child expected, per a verbal comment: ball (1)

17.3. Object somehow did not behave as child expected, per a verbal comment: bubble (1)

17.4. Object somehow did not behave as child expected, per a verbal comment: leaf (1)

17.5. Perceived unusually large object: ball (2)

17.6. Perceived unusually large object: flower (6)

17.7. Perceived unusually large object: leaf (1)

17.8. Specific ball could only be sorted by Andy, not the child (1)[4]

17.9. Unexpected that flowers could turn into balls (2)

17.10. Unexpected that same ball could change colour multiple times (1)

[4]Several activities created for ECHOES (but not used for most children) had "special" objects that only Andy could use, designated by appearance (e.g. a version of ball-sorting with the same goals, but special balls). This was supposed to foster co-operation with Andy, as the child could not complete the activity alone. Unfortunately, this did not go well in practice and these activities were not intended for use in the study; it may have been selected by mistake.

18. Violation of an internal ECHOES environment/activity pattern/rule (total = 15)

18.1. Absence of response or action from both partners when child completed an activity (1)[5]

18.2. All possible flower-picking answers were incorrect, activity seemed un-winnable (1)

18.3. Balls bounced/rolled out of side of screen and did not return (5)

18.4. Child and Andy threw balls at the same time and they hit each other (1)

18.5. No activity loaded, screen stayed dark (2)

18.6. Some balls were already in boxes at start of ball-sorting (1)

18.7. Unresponsive object became responsive again: ball (2)

18.8. Unresponsive object became responsive again: cloud (1)

18.9. Unresponsive object became responsive again: flower (1)

19. Violation of object expectations derived from out-with ECHOES (total = 4)

19.1. Balls rolled out through sides of basket (2)

19.2. Basket did not have a solid bottom, balls fell out (1)

19.3. Child grew a flower inside the basket (1)

E.2 Descriptions of aspect groups

NB: As all discrepant aspects are definitionally part of a pair with a child reaction, the current descriptions omit repetitive references to reactions. The presence of a reaction in relation to each aspect is assumed.

E.2.1 Novel aspect groups

1. Andy action resulted in ball-sorting activity reward: Child observed the sensory reward (bubbles, bees, or fireworks) that appeared when Andy correctly filled a box in the ball-sorting activity. This AG is separate from AG2 because of the element of reward for correct action, which is not present elsewhere in ECHOES. Discovering the rewards is not the same as discovering a new object action (or watching Andy do the

[5] Generally, completing an activity caused Andy to say "All done". The researcher usually repeated this or praised the child before suggesting a new activity. Here, for whatever reason, neither partner spoke until the child actively initiates—violating a session pattern of what happened when an activity finished.

same), and looking at the "effectiveness" of these as motivators may be of particular interest for future design.

2. Andy demonstrated novel action, object property, or relationship: The child observed Andy demonstrating a new action or sequence (e.g. jump to shake cloud, pick up a ball and put it in the box), demonstrating a new property of an object (the object itself may or may not already be familiar), or demonstrating a relationship between objects. There is some overlap inherent in this category. For example, demonstrating a new object property may have required a previously unseen action sequence. The duplicative labelling allows for the child's varying focus of attention: s/he may have paid more attention to the object, or to Andy himself. The key point of this category is that the child observed Andy rather than directly discovering the novel action.

3. Andy made a novel utterance: Andy made a previously unheard utterance, which may have been in response or reaction to a child's action, or for some other reason (e.g. introducing a new activity).

4. Child action resulted in ball-sorting activity reward: Child reacted to the sensory reward (bubbles, bees, or fireworks) that appeared when s/he finishes correctly filling a box in the ball-sorting activity. This AG is separate from AG6 because of the element of explicit, immediate reward for a correct action, which is not present elsewhere in ECHOES.

5. Child formed expectations about Andy: This aspect group included the child establishing expectations about "properties" of Andy that are independent of specific activities, including how and when he entered and exited (5.3), and the possibility of tickling him (5.1). Andy's novel language-related behaviour is treated separately, in AG3. This AG includes both interactive and observed aspects.

6. Child discovered novel object action, property, or relationship: Through direct interaction, the child discovered a new object action or property (e.g. pop bubble [6.5], pick up a ball and put it in the box [6.4]), or discovered a new relationship between objects (e.g flower pots change colour when stacked [6.9]). There is some overlap inherent in this category. For example, discovering a new object property (e.g. flowers turning into balls) sometimes also required a new action sequence. The duplicative labelling allows for the child's varying focus of attention: s/he may be paying most attention to the object, the action s/he is performing, the result of the action, and so forth. The key point of this category is that the child directly discovers the novel aspect him or herself, rather than observing Andy or the researcher's modelling.

7. Child was presented with a new scene: The child passively observed the entire configuration of the environment and perceived it as being new. S/he reacted to the "whole scene". The configuration may or may not have included individual familiar objects. Reactions to the whole scene generally occurred immediately after a new activity loaded, and before Andy was present.

8. Novel sound effect: Child perceived a sound effect as being new. Sound effects included any background or activity-related sounds that are not produced by the virtual character. The sound effect may be have resulted from direct child action (i.e. interactive) or have been unrelated to a child action (i.e. passive).

E.2.2 Surprising aspect groups

9. Andy interfered with the child's turn in an activity: The child attempted to take a turn in an activity (e.g. stacking the pots, sorting the balls) and Andy interfered with the action in some way. Almost all instances involve Andy "stealing" an object to take his own turn, a result of problems with the AI planner. This meant that objects appeared to magically jump away from the child's hand, and over to Andy. The child could not then take them back. This was a clear deviation from Andy's usual behaviour.

Along with the two "Child cannot perform expected object action..." groups (AG13, AG14), this is the only group that involves some kind of "barrier" to the child's play. Of those three, this is the only group where the barrier is directly related to the task/content. It is also the only one that involves interaction with the agent (though the word interaction is used loosely, and should not imply that all children saw Andy as social).

10. Andy was unresponsive to the child's communication: Andy had limited means for responding to child communication, and children sometimes reacted to this (e.g. Andy did not answer child's greeting so child initiates again, Andy does not react to poking and child comments about this).

This is the only group where the discrepant aspects can definitely be considered "social" in nature, which is of particular interest given the user group. For some children, instances from AG9 may have been perceived as Andy breaking social rules, but this is not very clear from the context.

11. Andy violated a pattern/rule regarding an activity's rules/goals: This group includes a range of Andy's "mistakes" or counter-productive actions in the activities (e.g. un-stack an already-stacked pot, throw ball and miss the cloud). These actions are mistaken or counter-productive relative to the rules, patterns, and/or goals already established within ECHOES.

12. Andy violated a pattern/rule regarding his own usual behaviour: Andy behaved differently than would have been predicted from his previously-established behaviour patterns (e.g. very delayed entrance, would not accept an offered object). Some instances were passively observed, and the child discovered others by attempting to interact with him. A key point about this AG is that it only includes behaviours that could be considered "general" properties or patterns for Andy, rather than those related to performing a specific activity (see AG11).

13. Child could not perform an expected object action due to touchscreen or gesture issue: The child attempted to use or manipulate an object in a particular way, but did not get the expected outcome due to an issue with the touch-screen hardware or with the child's physical input. As the AG only has non-events in it at this time, the actual outcome is *no outcome* (when some outcome, such as picking up and moving an object, was expected).

The title of the AG may seem vague, but this is because it is often impossible to tell from video data whether the child was touching the screen incorrectly (and thus the touch does not register), or whether the physical gesture is correct but there is some other issue. This AG also includes instances where the child began manipulating an object, but was unable to maintain the a continuous, correct touch over time and screen distance (thus dropping it repeatedly). In other cases, particularly the ball-sorting activity, children frequently struggled to move balls around other objects (causing them to ricochet away). It was generally impossible to "catch" a falling or ricocheting object, meaning the child generally had to start the whole action over. These were issues with the internal ECHOES physics engine rather than the touchscreen as a piece of hardware. While the physics engine played an important role in the overall function and experience of the environment, there were several ways in which it turned out to be inappropriate in combination with the large touch-screen and a user group with sometimes imprecise motor control.

14. Child could not perform an expected object action (non-touchscreen cause): The child attempted to use or manipulate an object in a particular way, but did not get the expected outcome due to some reason other than the touch screen. The majority of aspects in this group were the child being unable to perform actions that were disabled in particular activities, though there is one instance of the magic cloud failing to work for no discernible reason. Other aspects in this group included the child trying to perform actions that were not possible in ECHOES, but that s/he seemed to expect should work (hanging the flying leaves back on the tree, opening the gate pictured in the garden backdrop, pushing objects out of screen side to get rid of them). This latter collection of aspects was not placed in the group about violating expectations from out-with ECHOES (AG19) because the child's actions seemed to result from logical extrapolation from other ECHOES objects and actions, rather than directly applying knowledge from elsewhere.

15. Child perceived that an aspect was missing: Child perceived than a particular object should be present in the environment at that time, but was missing (e.g. cloud, flowers).

16. Mismatch between partner utterance and environment state: This aspect group includes instances of the child reacting because a social partner (human or Andy) has said something which does not match up with ECHOES. Most commonly, either the researcher or Andy says that Andy will take a turn (or do a specific action), but then there is a long pause where nothing happened. There are also instances where there was a direct contradiction between utterance and environment state at the point of the utterance (e.g. Andy says "all done" when activity is incomplete).

It is worth noting that this is a type of objective ECHOES error of which there numerous examples in the video samples. Due to delays in planning his behaviour, Andy's utterances frequently did not match up with the state of the environment, and declarations of his taking a turn or doing an action turned out to be untrue as often as not. However, it was relatively uncommon for children to perceptibly react to these language-based mismatches, though 5 of the 8 children did so at least once.

17. Unexpected or altered property of a familiar object: Child believed that a familiar (known) object was different than expected in some way. This could be in terms of its appearance (e.g. size, colour) or its properties (e.g. action on it produced a different result). An important difference between this group and AG18 is that these aspects are about differences that do not break any broader environmental rules. A perceived "giant ball" violates expectations only about the ball, not about the general way the environment operates.

18. Violation of an internal ECHOES environment/activity pattern/rule: This group covers instances where the child perceives that there has been a violation of some kind of general rule or pattern in how the ECHOES environment operates. These tended to be activity-independent rules, and frequently were due to objective errors on the part of the real-time AI planning, or in the original programming. For example, the one activity where balls bounced out of the sides of the screen violates the apparent

ECHOES rule that only Andy can move in and out of the screen sides (objects generally cannot; they rebound off). An instance where the ball-sorting activity loaded with some balls already in the boxes violated the established pattern of how the activity started on all previous plays.

This AG also contains several instances of what could be considered "micro-patterns" within interaction sessions. In the case of aspects 18.7-9, an object was apparently not responding to the child (for whatever reason) over a series of tries, and then eventually began working again. Children appear to have established new expectations over several unsuccessful tries (i.e. *this doesn't work any more*) which are subsequently violated when the object begins responding again.

19. Violation of object expectations derived from out-with ECHOES: This small AG includes the fairly rare instances in which two children appeared to react to violation of expectations derived from the physical world. All of the instances have to do with the basket, an ordinary-looking object which behaves in unexpected and inconsistent (but not "magical") ways. The basket can contain flowers and be moved with the flowers inside of it, but cannot contain balls: they fall out the bottom, and or roll out the apparently solid sides of the basket.

The presence of this AG highlights a phenomeon that might reasonably be expected, but is entirely missing from the dataset: instances of children reacting specifically to the magical aspects of the magic garden, and pointing out that things don't really work like that[6]. If not precisely "wrong", the objects and behaviours are unusual and unexpected. However, we do not see any of this—children seemed to accept the magic of ECHOES without comment, or not even see it as unusual. Given that all children in the study seemed already familiar with computer and television programmes, ECHOES is possibly *less* fantastical than other media they experience. There is also the issue of their developmental age. It is possible that they are not yet in a position, cognitively, to recognise some of the events as impossible, or at least unusual compared to the world.

[6]At least one participant in an ECHOES participatory design session objected to earlier versions of the system for exactly this reason. See Frauenberger et al. (2013).

Appendix F

Additional design recommendations from the ECHOES context

The high-level principles presented in Chapter 6 are principles for leveraging *discrepancy* as a part of a technology design, in a motivating but manageable way. These additional recommendations represent practical “lessons learned” from the ECHOES project. They make specific suggestions about design choices that work well—or poorly—for this type of context and participant group. They are *not* specific to the inclusion of discrepancy, but will inform the new designs detailed in the following chapter, and may be more generally useful to other ASC-tech researchers.

The recommendations are organised topically. Beyond these comments, the aspect groups and aspect type list (in Appendix E) are a valuable resource for choosing specific DDOs. In some cases, children’s subjective perceptions of ECHOES (such as objects that were not objectively giant, or not broken, or not missing) suggest objective alterations that could be made in a new design, and seem likely to attract child attention. Many of these ideas could be borrowed quite directly into a new design.

F.1 Action flow within an activity

In a design that seeks to promote communication, activities with smaller units of child action may better support that goal because they break up the game play into smaller pieces. They can provide interim events, achievements, or rewards as subjects of communication. Activities that enable prolonged actions (or action sequences) may better facilitate flow and engagement with the system, but provide fewer opportunities for communication and make may intra-session transitions difficult because children must first *dis*engage. It may be useful to have a more flow-like activity that could be used as a warm-up or reward activity, but structure the bulk of system content around more unitized activities.

F.2 Active and passive DDOs

The *VARIETY* principle already recommends that a future design include ample DDOs of both types, as child preferences seemed to vary (potentially related to self-inhibition ability). Another relevant factor here may be children's willingness to explore, once they have become confident about what to do in a particular game. Despite the apparent preference (numerically) for actively discovered discrepancies, some children were reluctant to explore in ECHOES, in the sense that they stuck narrowly to completing activities as these had been demonstrated. They did not touch around the screen to check if other things would respond, and did not engage in "what if..." type experimentation with known objects.

Making some DDOs immediately perceptible and available without active exploration (e.g. add or subtract a large object, change background soundtrack, character mistake) may increase the DDOs to which these children are exposed and could react. Alternately, DDOs could be created around objects and actions that the child will definitely use in order to meet the goal. However, in this case, it would be important to respect the *resolvability* principle.

F.3 Activity structure and goals

As discussed in Section 6.2.3.2, ECHOES participants struggled in the "free play" activity, which did not provide children with an obvious goal. The presence of many previously-seen objects (from activities with varying goals) definitely did *not* communicate to children that all goals were possible, and they could do as they liked. For these reasons, when designing for this same age and ability group it is recommended to create structured activities with goals, and to directly communicate the goal in some way (e.g. Malone and Lepper, 1987). Thus far, children seemed to find it more enjoyable and more manageable to explore *within* a structured activity ("What happens if I...") than to generate their own activity in free play. Designing DDOs will also likely be substantially easier in a structured activity with rules and goals. These provides a good way of setting up expectations and relationships that could later be altered as part of DDOs.

The ECHOES free-play also underscores that activities should probably focus on a single goal with no sub-goals, in the sense of children not needing to complete different *types* of actions or complete multiple steps to reach a goal. Even when it was explicitly explained or demonstrated, the current participants struggled to grasp how they could reach a goal through multiple steps (e.g. get a favourite object [ball] by using cloud to grow flower, then swiping to change it into a ball). Children had no trouble understanding repeated units of identical activity to reach a goal (e.g. grow multiple flowers).

Free-play type activities would still be possible and may sometimes be very useful and enjoyable for children. However, they should probably use completely separate objects or patterns than present in other activities and focus on sensory exploration or cause-effect relationships. Children may be less perplexed by goallessness if they do not have other cues leading them to expect a goal.

F.4 Adult social partner: roles and reporting

In ECHOES, the social partner was expected to be present as the child used the system, and their roles were defined beforehand at a high level. Going forward, an adult is assumed to be present in a receiving design context. The adult's role(s) should be specified in relation to the system throughout the design process, especially in terms of supporting and responding to the child. How will these functions be shared between the adult, the VC (if there is one) and other parts of the system? These roles then need to be tested and refined, alongside the technology proper. The manner and degree in which an adult supports a child user with ASC is not "supplementary online material", but a basic part of considering what a technology does, and for what abilities and contexts it might be suitable. It is strongly advised to collect data in such a way as to capture supporting adults (as with the video positioning in ECHOES) and then to also report adult roles and actions as a part of reporting on a design and its evaluation. This type of deliberate design and reporting of adult roles allows a researcher—and readers—to better reflect on the effectiveness and apparent value added by each component of technology use in context (adult, technology, other supplementary supports).

F.5 Touch-screen interface and actions

All ECHOES participants, including those developmentally and chronologically oldest, had some trouble with the touch screen interface. Any designer using a touch-screen with this group should plan for problems and how to diffuse them. In some respects, the most difficult activities for children to complete were those with the most difficult touch-screen actions. Comments on specific actions and issues are as follows:

- Combination of touch screen and environment physics: Physics were important to ECHOES, as many objects bounced, swayed, and dropped. These behaviours were a key part of the design. However, they were also a main source of difficulty. Children had a very hard time lifting and dragging objects (see below), and losing hold of objects when these collided with their neighbours. Minimising the physics in future activities might cut out many of the touch screen problems. An alternative might be to drastically *slow* the physics, so children would have a chance to catch falling or ricocheting objects, and continue their intended action.

- Tapping or single-touches: Overall this was the most successful type of action, and the easiest for children to grasp.

- Dragging objects: When combined with physics and the large screen size (42”), this was a major source of frustration and delay, especially in Ball Sorting and Ball Throwing. Children struggled to maintain consistent pressure (at least on a vertical screen) and, when they briefly released an object it would drop. They then needed to start their action over. It is unclear whether a smaller screen size, with shorter dragging distances, might ameliorate some of this problem.

- Long (held) touch without dragging: Difficult for some children to grasp initially, but relatively easy thereafter. This is what children did when using the magic cloud for long periods of activity.

- Multi-touch screen: Children rarely used both hands in a deliberate way, but sometimes children unthinkingly leaned on the screen with one hand and moved objects with the other. This would not be possible on a single-touch screen, and would likely cause an error.

- Multiple-finger actions (e.g. pinch): Not used in ECHOES; thought to be too difficult for this age group.

F.6 Virtual character (To VC or not VC)

Within ECHOES, the VC Andy played a number of useful roles, and all participating children attended to him and responded to him to some degree. In relation to DR pairs, several of his roles seem particularly relevant:

As an actor in the environment Introducing a character gives the opportunity to establish a large set of additional, inherently novel behaviours, patterns and rules—all of which can then be changed or added to as DDOs. Even if children do not think of the VC in a social way, he can still be a special mobile object that establishes and violates patterns. Adding a character dramatically expands the designer's options.

As a demonstrator or teacher Andy's primary responsibility for demonstrating activities and giving directions may have contributed to the *adult's* successfulness as a social partner. When Andy gives directions, the adult does not need to, separating a teaching-type role from an affective, responsive "partner" role. Not having the adult demonstrate and give directions (i.e. already be an expert in the environment) may have been important in increasing children's sense of ownership over the environment, and also making parts of it more worth sharing with the adult, who is *not obviously knowledgeable* about the activities or technology.

As a recipient of directions and help Several children in ECHOES (particularly Odell and Kalil) tried to help Andy and give him directions when things went wrong, allowing them to step into a knowledgeable, helping role. This may be a very positive, confidence building experience for some children. A future design might deliberately try to provide helping opportunities of this type.

In ECHOES, Andy's behaviour was deliberately and reactively planned by a planning engine (see Chapter 5). He planned and executed different actions depending on what the child did. Due to the delay caused by planning and execution, Andy was always reacting to what the child did some seconds before; this was a source of many discrepancies in itself. A future VC could probably be much simpler and more scripted, but still have the same benefits and play the same roles as an AI-driven character. Relevant VC reactions to the child could be easily provided by hot-keys or a GUI menu for the adult partner (e.g. pressing a key to make the VC repeat instructions).

Appendix G

Chapter 7 supplement

G.1 GameSalad game editor

The logic of development in GameSalad is largely that of object-oriented programming, recast in the language of theatrical performance. Within a game project file, the user may create one or more *scenes*, each of which may function as a menu, transition, action sequence, or activity. The user then defines the action rules and attributes of *actors* at class (*prototype*) and object (*instance*) level. These might be characters, objects, visual effects, timers, and so forth. Actors do not need to be re-defined for each scene within the same game project. Both scenes and actors may have variables of different types. A notable disadvantage of GameSalad is that it is extremely cumbersome for actors to "talk" to each other, because they cannot directly access or change each other's variables—only their own variables, and game variables. To coordinate the actions of different actors requires repeatedly storing, changing, and accessing values in many separate game variables. Convoluted variables aside, the object-oriented style of GameSalad was very useful when it came to implementing DDOs, because it meant that it was easy to create special instances of objects with slightly different behaviours than the main class.

The editor (Figure G.1) allows the user to directly move and re-size actors on the *stage*, which is very useful for designing a primarily visual product such as this. Game behaviours and timing can also be tested by playing in a preview mode, that has all the functionality that a user would see. Incorrectly defined rules and behaviours simply will not work or will work incompletely, but the rest of the game is intact. This proved to be extremely valuable for debugging, and fine-tuning how the games behaved.

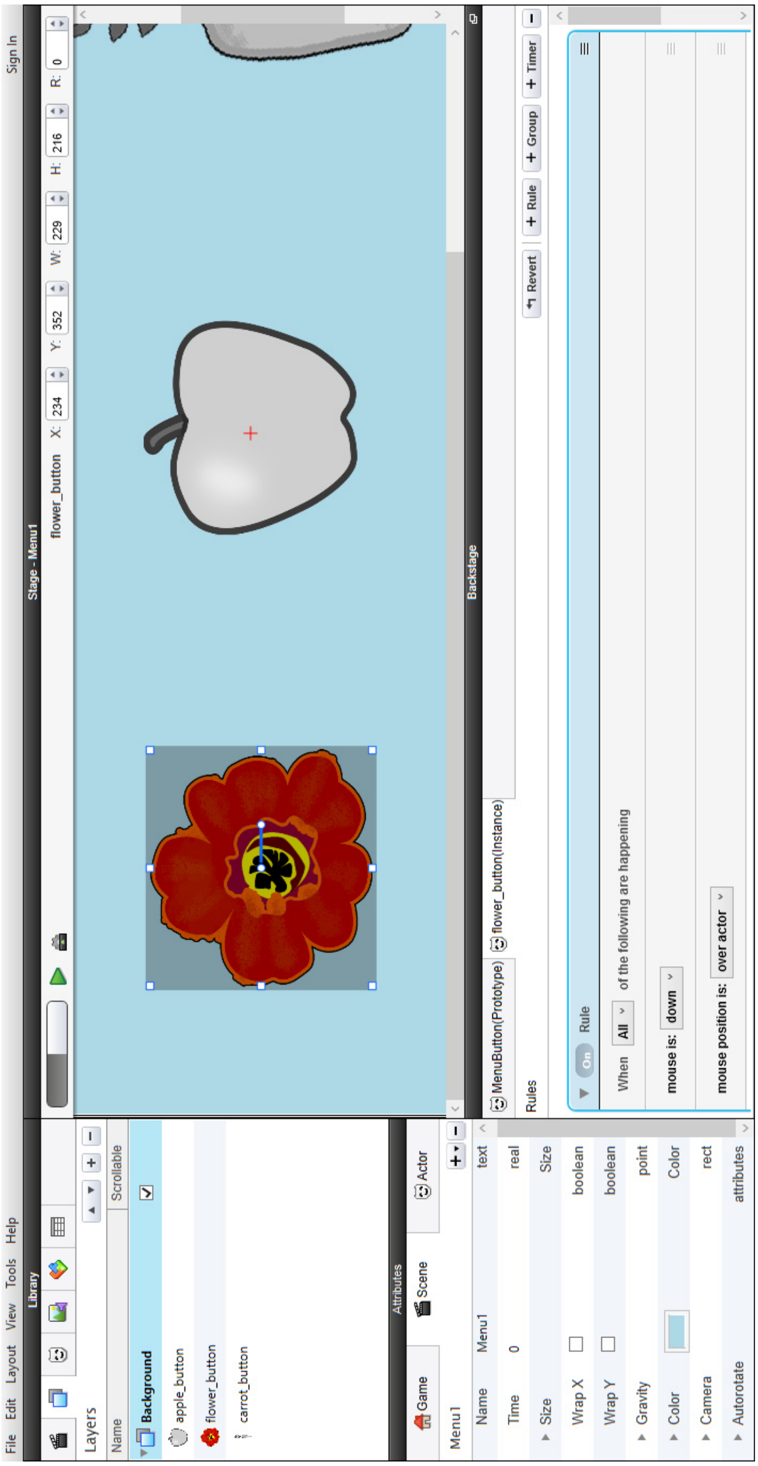


Figure G.1: Sceenshot of the GameSalad editor, showing the menu scene. Left: tabbed panels for managing scenes and actors. Bottom: *backstage* panel for defining and managing actor behaviours. Can be resized or hidden. Center: *stage* with layout of the menu scene, with the *flower button* actor currently selected. A child presses the flower button to go to the flower game, implemented as a separate scene.

Sound effect	Source	Use permissions
Daydreaming	http://soundbible.com/2030-Daydreaming.html	Public Domain
Pop Cork	http://soundbible.com/533-Pop-Cork.html	CC Attribution 3.0
Fireworks and Crowd	http://soundbible.com/712-Fireworks-And-Crowd.html	CC Attribution 3.0
Bike Horn	http://soundbible.com/1446-Bike-Horn.html	CC Sampling Plus 1.0
Butterfly	http://soundbible.com/1322-Butterfly.html	CC Attribution 3.0
Blop	http://soundbible.com/2067-Blop.html	CC Attribution 3.0
Magic Wand Noise	http://soundbible.com/474-Magic-Wand-Noise.html	CC Attribution 3.0
Thunder	http://soundbible.com/1907-Thunder.html	CC Attribution 3.0
Electrical Sweep	http://soundbible.com/1795-Electrical-Sweep.html	Public Domain
Winning Triumphal Fanfare	http://soundbible.com/1823-Winning-Triumphal-Fanfare.html	CC Attribution 3.0
Magical	http://soundbible.com/1088-Magical.html	CC Attribution 3.0
Eating Chips	http://soundbible.com/402-Eating-Chips.html	CC Attribution 3.0
Rockslide Small	http://soundbible.com/1987-Rockslide-Small.html	CC Attribution 3.0
Gum Bubble Pop	http://soundbible.com/1982-Gum-Bubble-Pop.html	CC Attribution 3.0

Table G.1: Non-voice sound effects used in *Andy's Garden* game suite. Sounds are named per their original sources, rather than referring to how they were used in the games.

G.2 Sound attributions

Non-voice sound effects were sourced from SoundBible.com, a site specialising in royalty-free sound effects. Those sounds that were not in the public domain were re-used and/or modified under Creative Commons licensing (see Table G.1). The voice of Andy was recorded specifically for the game, and was performed by Leon Blake. A full list of the dialogue used in the game is listed in the following section.

G.3 Virtual character dialogue

The following list gives all character dialogue in *Andy's Garden*, grouped by topic or function within the game. In some cases, these phrases have been re-combined in different ways (e.g. "Let's tidy up the apples. Your turn!"). The components are listed here, rather than all the possible combinations.

All editing for effects and Andy's voice was done in Audacity, a free, open-source editing and recording programme (Windows version 2.0.6; Ash et al., 2014).

Starting the game

Hello

Hi, I’m Andy.

Let’s play in the garden.

Will you play with me?

Andy gives directions and explanations

Let’s tidy up the apples.

Put an apple in the tube.

Look, a cloud!

Let’s make it rain.

Let’s grow some flowers.

Let’s grow some carrots.

Can you pick a carrot?

Your turn

My turn

Look! (as in “look at that”)

Andy tells the player that she/he is doing a good job

Thank you

Well done!

Good job

You did it!

Wow!

Andy comments on interesting things in the garden

Mmm, a carrot!

Mmm, a strawberry!

I like apples.

Look, a red apple!

Look, a yellow apple!

It’s a red flower!

It’s a yellow flower!

It’s a pink flower!

It's a purple flower!
It's raining on me!

Andy explains that a game is finished

All done!
Finished
You tidied up all the apples!
My mum will make me an apple pie.
You grew all the flowers!
These flowers are for my mum.
You grew all the carrots!
Thank you for helping me.

Saying goodbye

I need to go home now.
Let's play again soon!
Bye!

G.4 DDO descriptions

This section gives a very brief description of each DDO included in the altered versions of Andy's Garden. They are presented in the same order as in Table 7.3. This groups DDOs by their position within an activity (start, main game play, or end), and list them alphabetically within such groups. Several DDOs are discussed in more detail in-text, marked with "**".

G.4.1 DDOs during the activity start

Changed growing behaviour: Plant partly grown at start A flower or carrot is partly grown at the start of the activity and can now be completed more quickly.

Cloud missing at activity start** When a growing activity loads, the cloud is absent instead of being in the upper left of the screen. After some seconds (a noticeable pause), this DDO self-resolves when the cloud glides onto the screen to its usual start position.

Fence removed from garden White picket fence removed from background in growing scene.

Planter colour changed One planter box changed from dark brown to light brown in growing scene.

Snail added to garden A small green snail appears in the bottom right of the garden, by the corner of the planter box. When touched, the snail makes a popping noise and extends his head and tail out of his shell. When touched again he retracts. After a maximum number of touches in the same activity, the snail stops responding.

Some apples do not fall from tree Apples usually fall off the apple tree as soon as the apple game loads. Now, two remain on the tree until they are touched to "pick" them. They can also vanish off the tree if Andy selects one of them with which to take his turn.

Some apples giant-sized When the apple game loads, two apples (one each red and yellow) are noticeably larger than all the others. All of their other properties remain the same.

G.4.2 DDOs during main game play

Andy apple mistake: throws to wrong tube** Andy throws an apple on a trajectory such that he misses the right colour tube, and throws it to the other tube or the white tube. He is set to miss his first turn of the activity, and may randomly select a wrong turn thereafter (i.e. choice of trajectory is now randomised, rather than being set by apple colour).

Andy apple mistake: throws and misses tube** Andy throws an apple on a trajectory such that it falls short of the tubes apparatus entirely and falls to the ground. Implemented as above.

Andy mis-statement** Andy says something incorrect (e.g. says that a yellow flower is a red flower, says that a strawberry is a carrot) or totally irrelevant (e.g. he gets rained on and says he likes apples). He does not introduce any new phrases, he only incorrectly applies phrases used elsewhere (to ensure the phrases themselves are non-novel). This DDO is set to happen in response to specific events (e.g. completing a particular flower) rather than choosing between phrases and randomly being wrong some of the time.

Andy new feedback: You did it!** New feedback phrase to the set from which Andy's feedback phrase is randomly chosen. As this makes only four phrase options in total, the child is very likely to hear this several times within each activity, unless s/he plays very quickly.

Butterfly flies across screen An orange butterfly flaps across the screen, accompanied by a wing-flapping sound effect (starts 3 seconds before the butterfly comes on screen). The butterfly moves fairly quickly and is all the way across in about 8 seconds. It can also be triggered by a researcher hot key for an "on demand" DDO.

Carrot turns into strawberry when picked** Self-explanatory. Uses usual popping sound effect.

Changed growing behaviour: Plant grows in 1 rain Only one "unit" of rain (i.e. graphic of a group of rain drops, equivalent of about 1 second of cloud-pressing) is required to grow a plant to completion.

Cloud sound changed: Thunder Thunder sound effect replaces usual rain sound effect when cloud is pressed. Lasts entire activity.

Cloud sound missing Rain sound is missing on some touches. Is intermittent; applies to first time cloud is used in the activity and selected other uses. The cloud actor has a "touch counter" and triggers this sound when it has a certain number of touches (e.g. 1, 2, 5, 9).

Fanfare changed to novel sound The trumpet fanfare for a completed plant is replaced with a new sound effect. In one activity this is a honking, "bike horn" effect. In another it is a chime or xylophone-like effect.

Flowers: All different All flowers are different from their normal colours. All tulips are pink-and-white instead of purple, and all poppies are yellow instead of red.

Flowers: One different A yellow poppy is substituted for one of the red poppies, always in the same position.

Giant raindrops Usual raindrop graphic is substituted for a small cluster of giant rain drops. These have the same properties as normal raindrops. Implementation same as *Cloud sound missing*, above.

Some carrots giant-sized when picked Self-explanatory. Affects two carrots in the same activity, always in the same positions.

Tube sound changed to novel sound For the second and fourth apples sorted, the tube changes it usual sound to a crumpled-paper-type sound (very different than the usual zippy, slippery tube sounds).

Tube sound missing No sound effect while an apple "goes through" a tube. There is a fanfare as usual then the apple appears in the basket a few seconds later.

White tube makes 2 apples** Tube produces two apples of an opposite colour to the input apple. Tube is set to have this effect the first time it is used in an activity, but then returns to its usual effect (produce one apple of the opposite colour). The tube re-sets after a few minutes, and will produce the two apple effect again if the child is still playing.

White tube makes strawberry** Tube turns either colour apple into a strawberry. Implementation same as previous.

G.4.3 DDOs at the end of an activity

Andy changes end-of-activity utterance** After saying the activity is finished, Andy replaces his usual phrase with a new phrase that introduces some additional narrative. In the apple game, he says "I like apples. My mum will make me an apple pie!" and in flowers he says "These flowers are for my mum. Thank you for helping me!". After a brief pause the activities exit back to the menu as usual.

Andy disappears at end of activity After all of his end-of-activity utterances, Andy suddenly disappears with a "pop" sound effect.

Fireworks reward sound changed: magic noise The reward animation is as usual, but is accompanied by a tinkly "magical" noise rather than by fireworks and cheering. The new sound effect is noticeably quieter.

Rainbow appears at activity end When an activity is completed and the reward effect has finished, a brightly-coloured rainbow appears centre screen, above Andy. It is not interactive and its appearance is not accompanied by any sound effects.

Appendix H

Ethics and information forms for Study 2

H.1 Information and consent forms for games testing with typically developing children

Testing touch-screen games for children with autism

Information sheet for parents and guardians

This information sheet is for parents and guardians and explains a research project in which we would like your child to participate as a "games tester". It gives information about the project in the form of questions you might have and their answers. If you have further questions, we are happy to discuss them and give you more information.

The researchers on this project and their contact details are as follows:

Ms. Alyssa Alcorn, lead researcher (aalcorn@ed.ac.uk; *mobile number redacted*)

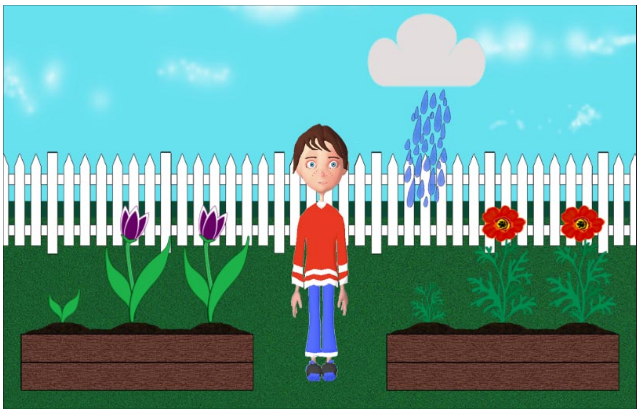
Dr. Helen Pain, research advisor (helen@inf.ed.ac.uk)

University of Edinburgh, School of Informatics
Informatics Forum
10 Crichton Street
Edinburgh
EH8 9AB

Please return the parent form to Alyssa or Helen if you give permission for your child to participate in the research project.

An overview of this project, and how you can help

What is the project goal? This testing session is part of a research project that tries to better understand how using novel and surprising elements in computer games might motivate children with autism to communicate with the people around them more often. We will do this by observing childrens' interaction with a new set of touch-screen games that include a range of novel and surprising elements. *For more information about the overall research project, please see page 4.*



How can my child help? To learn more about using novel and surprising game elements to motivate communication, we have built new touch-screen games *(see above right)* related to the original ECHOES *(see above left)*. These deliberately incorporate a wide range of novel and surprising events, object, effects, and character behaviours.

Your child can help us as a "games tester" for these new activities, in order to ensure that they are easily usable by (and interesting to) young children. Prior to testing any software with children on the autism spectrum, we always test early versions with typically developing children of a similar age to the autism target group. This gives us a reasonably good idea of how children with autism will be able to use the system, both in terms of physically using the touch screen, and understanding the game's goals and instructions. Typically developing children also frequently have stronger verbal language skills than their peers with autism, and can tell us more clearly what they like or dislike about the game, and make suggestions for how it could be different. They also tend to have greater tolerance when things go wrong with early, sometimes buggy, software versions.

The games are set in a cartoon-like "Magic Garden" with a character called Andy, who is a helper for the child. They are exploratory and focus on cause-and-effect play, though all games have a task that provides some goals and structure. For example, in the Apple Sorting game *(right)*, the child and Andy put away red and yellow apples into the tubes of the same colour, and are rewarded by sound effects. The games are meant to be fun and non-competitive. They do not have any points, levels, or timers, and there is no winning or losing. The goal is for the child to follow his or her interests within the Magic Garden and to feel successful by participating. No reading is required.

What happens during a games session? At a time we have agreed with you, both you and your child visit our office at the University of Edinburgh Informatics Forum. After explaining again to the child what will be involved in being a games tester and checking that it is OK with him or her to video the session, the child will get to play the games for approximately 20 minutes (or until s/he wants to be finished, whichever is first). It is fine for the child to take breaks to rest, have a snack, etc. Parents are welcome to be present during the game session, but unless the child appears anxious we usually ask that they sit on the other side of the office, or outside (the office has a glass wall). It can be distracting for the child when the parent is right next to him or her.

Immediately after the session, the child will get a coloured, laminated certificate saying that he or she has helped us as a games tester. In the past, these have been very popular with children at nursery or school show-and-tell!

What else does taking part in the study involve? We will want to take a video of your child testing the games. These videos are our main source of information for this study, and an important part of our methods overall. It means that researchers can watch the videos later (multiple times if needed) to explore how a child used the game, count certain types of events or child reactions (most of which happen very quickly), and potentially discover new, unexpected patterns in the child-technology or child-adult interactions.

Especially when children may frequently request help or otherwise be interacting with the researcher, it is NOT possible for us to replace video with extensive note-taking. If you prefer that we do not use videos or pictures of your child for publications, presentations, or teaching purposes, you can indicate this on the permission form. In that case, the video would be seen only by the research team.

Additional study information

What happens when the research is over? The researchers will use the results of games testing to improve the games, for example changing the character's dialogue to be simpler or adjusting the timing between events. After these changes, the software will be ready for in-school studies with young children on the autism spectrum.

We will share our overall project results with other researchers working on technologies for children with autism by publishing papers in scientific journals and giving presentations at conferences. This often also includes reporting on the testing process and giving specific examples of children's comments or interactions, and how these led to software changes. We may also discuss the "games testing" aspects of the research when teaching university students who are studying technologies for children. We sometimes use videos during presentations or teaching, because it can be hard to explain how a child interacts with a game but very easy to show it. We never share childrens' names, schools, or other personal information as a part of publications, conferences or teaching.

How will personal information be protected? Confidentiality is extremely important to us. Videos and other information (such as forms with children's names) will be stored safely on password-protected computers or in locked cabinets. Access will be limited to the people involved in this research: Alyssa Alcorn and her research advisors. Videos and other information will be identified only by participant codes or pseudonyms, and will be completely separated from identifying information (such as name or birth date). In publications and research presentations, we refer to children by codes or pseudonyms.

Can we have a copy of the games? The games your child will be testing (in an unfinished version) are research software and are designed to help answer specific questions. They do not have the same type of functionality, amount of content, or style of documentation that you might expect from a commercial game. However, once the final study is over the games will still be made publicly available online, for free unlimited use. This should be in early 2015. You will be sent a link and starting instructions; all you will need in order to play is a computer with internet access.

Who paid for this research? This study is part of the PhD work for the main researcher, Alyssa Alcorn. It is indirectly paid for by University of Edinburgh-funded PhD scholarships, which are not attached to a specific project or to any outcomes of that project. Conducting this research brings no additional financial benefit to the researchers or to the university.

Would you and your child like to participate?

We ask parents to read this information sheet and to decide whether they think it is a good idea for their child to participate as a games tester. Parents may also indicate their choices about how project information is used.

If you say "yes" on this permission form, we would ask you to then explain the games tester role to your child, and ask him or her if s/he wants to help. We have included a children's information sheet with a simplified explanation of the study, and this may help you in discussing the games testing with your child. Please do ask your child what s/he wants to do. We feel strongly that children should be given a real choice about whether to participate in research--or not.

Once again, participating as a games tester is *completely voluntary* and you and your child are under no obligation to take part. Even if you say yes now, you may later withdraw your child *at any time and for any reason* by contacting the researchers. Your child may also withdraw at any time during the testing session by saying that s/he does not want to be a games tester any more.

Thank you for taking the time to read this.

If you would like to know more about this research and/or if you have questions, please feel free to contact Ms. Alyssa Alcorn from the University of Edinburgh School of Informatics, (aalcorn@inf.ed.ac.uk) or on *mobile number redacted.*

More information: Why research surprising and novel elements in games for children with autism?

The main researcher (Alyssa Alcorn) is conducting a PhD project on novel and surprising elements in interactive games intended for young children with autism, and how these can sometimes motivate them to practice their communication skills. Social communication skills are one of the main areas affected by autism, a group of developmental disorders.

The current project came about because a previous project (a touch-screen game called ECHOES), made some unexpected observations about things that may motivate children with autism to initiate new communications. As with any experimental piece of software, ECHOES had some errors and malfunctions. We had been concerned that it might upset children with autism when the objects and the animated character, Andy, started breaking "the rules" of the game. However, most children appeared very interested in these "surprising" game events, and also in new objects and character behaviours. They seemed to find them fun and entertaining.

Children often turned to adults in the room to laugh and smile about surprising and novel events, or to comment and ask questions on them. Children starting new, spontaneous communications was a very positive but completely unexpected thing to observe, because this type of behaviour is relatively rare in children with ASC (compared to typically developing children). However, it is a type of behaviour that is important for long-term development and many autism support programmes try to teach or increase this kind of communication.

We currently know very little about *why* novel and surprising parts of the game seemed to be so interesting or "worth communicating about". Analysing the ECHOES data has allowed us to develop some possible explanations about why these things were interesting, and why children were not upset by them. However, these ideas need to be tested by building a new game and observing children's behaviour as they play with it. This is the goal of the current study.

New computer games to help children

This page is about some people who want to ask for your help. They are trying to answer a question about children and computers. It says who they are, and what they are doing.

You can help them by playing a new game and saying what you think about it. You can decide if you want to say “yes” or “no” to helping.

Who is asking for my help?

This is Alyssa and Helen. Their job is to learn about how children use computers, and how to make computer games that can help children. They will ask you to help them by playing a new computer game.

Alyssa

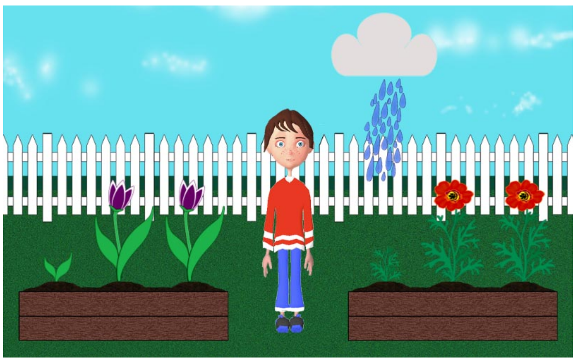
Helen

Why are you asking for my help?

It is hard for some children to talk to other people. Alyssa and Helen have a question about this. “Can playing special computer games can help these children to talk to other people more?”

Alyssa and Helen will ask children like you to help them find the answer. Children can help by being “games testers” and playing a new computer game.

This game is about a Magic Garden. It is very new. No children have played it yet! There are lots of things to do in the garden, but Alyssa and Helen do not know which things children will want to talk about. Here is a picture of the Magic Garden:



Alyssa and Helen will ask different children to be “games testers”. They can play the game and say what they think. This game is just for fun. It does not have any levels or points. No one wins or loses. Some children may think it is easy. Other children may think it is hard. That is OK. All children can help. Alyssa and Helen will learn when they see children play the game.

This page says how you can help as a games tester. Then you can decide if you want to say "yes" or "no" to helping.

What will happen if I say "yes" to helping?

Alyssa and Helen will ask you to visit the place where they work. Your mum or dad will come too. When your mum or dad says it is OK, you will get to play the new game on a big touch-screen. Alyssa and Helen think that you will like playing this game.

They will ask some questions about things you liked or did not like. Anything you want to say about the game can help us. Children know much more than grown-ups do about what other children like. There are no right or wrong answers. If you don't want to give an answer, that is OK too.

You can tell Alyssa and Helen if you want to stop playing the game. You do not have to tell them why. Please tell them if you need the toilet, or if you want to take a break. You can also say you do not want to be a games tester any more, and that is OK. Alyssa and Helen will always listen to you.

Alyssa and Helen will have a video camera. They will ask if it is OK to make a video of you playing the game. This is because it is too hard for Alyssa and Helen to write down everything that happens. They will watch the video later to help them understand what you said, and what happened when you played the game.

What will happen after I play the game?

Alyssa and Helen will learn a lot about children and computers when we watch children play the game. They think this will help them to make new games that can help children to talk more. They will write about what they have learned. When Alyssa and Helen talk to other people who make games, they will say what they learned. Sometimes Alyssa and Helen will show people videos of children playing or talking about the game. Those things are easy to show in a video, but hard to say in words!

When Alyssa and Helen are all finished, they will send another page like this one to you and your parents. They will say "Thank you!" for your help. The page will say what you did to help them. It will explain what they learned about children and computers.

Your mum or dad said it is OK for you to help us.

Do you want to be a games tester? You can say "yes" or "no". It is OK to say "no". It will not hurt Alyssa and Helen's feelings.

Do you want to ask a question about being a games tester?

It is OK to have more questions. You can ask Alyssa as many questions as you want about being a games tester. Ask your mum or dad to help you call her on the phone or write an e-mail with your question.

Mobile phone number: *mobile number redacted*
E-mail address: aalcorn@ed.ac.uk

Research permission form (for parents)

	Please circle
Have you read the information sheet(s)?	YES / NO
Have you received enough information about the study?	YES / NO
Do you understand that participation is completely voluntary and your child can leave the games testing session (or any individual activity) at any time, without having to give a reason?	YES / NO

Please sign this page to indicate that you understand and accept the conditions of this study, including video recording of the testing session. By signing, you agree that the researchers may explain the study to your child and invite him or her to take part as a games tester.

With reference to further anonymous use of photographic or video data, please circle yes or no in response to the following:

> I AGREE that short videos/images of my child can be used as examples in documents and presentations for research and/or teaching purposes.
>
> YES / NO

If you DO NOT wish to give permission for this study, you do not need to return this page or take any further action. We will not ask your child to participate.

Full name of participating child: ____

Child's date of birth (DD/MM/YYYY): _______/_______/_______

Your relationship to the child: ____

Your name (please print clearly):____

Contact telephone number: ____

e-mail address: ____

Best time and method to reach you? ____

Signature: ____

Date: _______ / ______ /______

Child Consent Form

To be used as a guide for securing consent or refusal, after the child has had a chance to get information about the study. The child may mark (or be helped to mark) this form, or the child's consent/refusal may be video-recorded.

I can choose to be a games tester.

I do not have to help if I don't want to.

I can decide to stop playing or take a break.

I do not have to say why.

It is OK if I change my mind later, and say I do not want to be a games tester any more.

It is OK if some parts of the game are hard for me!

Anything I can do is helpful.

Do you want to be a games tester? **YES ☐ NO☐**

Alyssa and Helen will have a video camera to record what happens when I play the game.

Alyssa and Helen will look at this video later.

They will show it to other people who make games for children.

Is it OK to take a video? **YES ☐ NO☐**

Write your name: ___________________________________

THANK YOU!

H.2 Information and consent forms for school study with children with ASC

Touch-screen games for children with autism: A research project

Information sheet for parents and guardians

This information sheet is for parents and guardians and explains a research project at Kaimes School in which we would like your child to participate. It gives information about the project in the form of questions you might have and their answers. If you have further questions, we are happy to discuss them and give you more information.

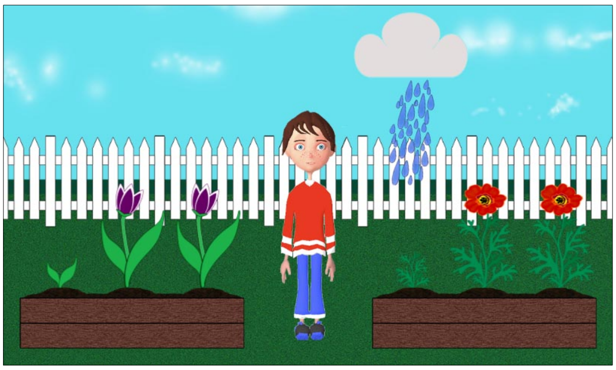
The researchers on this project and their contact details are as follows:
Ms. Alyssa Alcorn, lead researcher (aalcorn@ed.ac.uk; *mobile number redacted*)
Prof. Helen Pain, research advisor (helen@inf.ed.ac.uk)

University of Edinburgh, School of Informatics
Informatics Forum
10 Crichton Street
Edinburgh
EH8 9AB

Please return the parent form to your child's teacher if you give permission for your child to participate in the project.

An overview of this project, and how you can help

What is the project goal? This project tries to better understand how using novel and surprising elements in computer games might motivate children with autism to communicate with the people around them more often. We will do this by observing childrens' interaction with a new set of touch-screen games that include a range of novel and surprising elements.



Why do this project? Our previous work with a touch-screen game called ECHOES made some unexpected observations about things that may motivate children with autism to initiate communications. As with any new game, ECHOES had some errors and malfunctions. We had been concerned that it might upset children with autism when the objects and the animated character, Andy, broke "the rules" of the game. However, most children appeared interested in these surprising elements, and also in exploring new objects and character behaviours. They seemed to find them fun and entertaining.

Children often turned to adults in the room to laugh and smile about surprising and novel events, or to comment and ask questions on them. Children starting new, spontaneous communications was a very positive but completely unexpected thing to observe. As you may already know, this is the type of behaviour that is relatively rare in autism, but very important for long-term development. Many autism support programmes try to teach or increase this kind of communication.

We currently know very little about *why* novel and surprising game elements seemed to be so interesting to children or why they were "worth communicating about". Analysing information from ECHOES has allowed us to develop possible explanations about why these things were interesting, and why children were not upset by them. However, these explanations need to be tested by building a new game and observing children's behaviour as they play with it.

How can my child help? To learn more about using novel and surprising game events to motivate communication, we have built new touch-screen games related to the original ECHOES. The games are set in a cartoon-like "Magic Garden" with a character called Andy, who is a helper for the child. They are exploratory and focus on cause-and-effect play, though all games have a task that provides some goals and structure. For example, in the Apple Sorting game *(right)*, the child and Andy put away red and yellow apples into the tubes of the same colour, and are rewarded by sound effects. The games are meant to be fun and non-competitive. They do not have any points, levels, or timers, and there is no winning or losing. The goal is for your child to follow his or her interests within the Magic Garden and to feel successful by participating. No reading is required.

We would like your child to try playing these activities, as a "games tester". We will try to find out what types of things in the game interest children most, and are worth laughing at, asking questions about, or otherwise sharing with another person. Most of our information will come from observing their play, but we may also ask a small number of questions (either verbally or supported by signs and pictures, as appropriate for your child).

What happens during a games session? At a time agreed with your child's teacher, we will walk him/her from the classroom to the quiet space where the computer and touch screen are set up. In some cases, a staff member may accompany your child if the teacher thinks that this will help him or her to be more comfortable. Your child will get to play the games for 15-20 minutes, or until they are tired (whichever is first). We hope to have 3-4 games sessions with each child, spread over two weeks of school.

Please note that we always make it our policy to visit schools beforehand, and make sure that researchers and participating children have a chance to become familiar with one another during a full-class activity. We always do this *before* games sessions begin, which had been very important for helping children to be more comfortable.

What else does taking part in the study involve?

Recording games sessions: We will want to take videos of your child using the games. These videos are our main source of information for this study, and an important part of our methods overall. It means that researchers can watch the videos later (multiple times if needed) to explore how a child used the game, count certain types of events or child reactions (most of which happen very quickly), and potentially discover new, unexpected patterns in the child-technology or child-adult interactions.

It is NOT possible for us to replace video with extensive note-taking, especially as children may frequently request help or otherwise interact with the researcher. If you prefer that we do not use videos or pictures of your child for publications, presentations, or teaching purposes, you can indicate this on the permission form. In that case, the video would be seen only by the research team during the analysis. If you are not comfortable with your child being videoed at all, then your child should not participate in this particular study. However, there may be other opportunities for them to get involved with autism and technology research and we are happy to discuss these when they arise.

Background information: The researchers will not know your child well, and cannot compare what your child says and does during the game sessions to what he or she is "usually like". We will want to collect background information in order to get an overall picture of your child's language comprehension and other communication skills. This is important because we have designed the games with a target group of children in mind. We want to know if we have succeeded in creating a game that is usable, understandable, and fun for that group. We also want to know if children with different abilities use and understand the game in the same way.

We plan to use two standard measures: the British Picture Vocabulary Scale (BPVS), and the Social Communication Questionnaire (SCQ). The BPVS is a widely-used test of a child's verbal language understanding. It takes 5-8 minutes to administer at the school. If the school's Speech and Language Therapist already uses the BPVS and your child has a recent score, we will use this existing score to avoid repeat testing. We will also ask you to complete the SCQ questionnaire for parents/guardians, which has 40 questions about your child's communication behaviour. It will take you less than 10 minutes to complete.

Additional study information

Will this project teach my child new skills? This project is not a type of therapy or an intervention. We will not be teaching children new skills or improving existing skills. The information we learn from this project may be used in future games that could be tools for teaching or improving communication skills for children with autism.

What happens when the project is over? After the Kaimes School study has finished and the researchers have analysed the information we collected, we will share our results. This includes descriptions of how children interacted with the game, and counts of how often children reacted to certain parts of the game or used different communication behaviours. We will use this information to help us make recommendations for what to include (or not include) in future games that are meant to help children initiate communication more frequently.

We will share our results with other researchers who create technologies for children with autism by publishing papers in scientific journals, and presenting at conferences. We may also discuss this research when teaching university students who study technologies for children. We often use videos during presentations or teaching, because it can be hard to explain how a child interacts with a game but very easy to show it. We never share children's names, schools, or other personal information.

It is important to us that schools, teachers, child games testers, and their parents also hear what we found out during the project. It takes several months to analyse data, but after that time we may send home video clips or information pages like this one, or may come back to the school to talk about the project results.

How will personal information be protected? Confidentiality is extremely important to us. Videos and other information (such as forms with children's names) will be stored safely on password-protected computers or in locked cabinets. Access will be limited to the people involved in this research: Alyssa Alcorn and her research advisors. Videos and other information will be identified only by participant codes or pseudonyms, and will be separated from identifying information (such as name or birth date). In publications and research presentations, we refer to children by codes or pseudonyms and do not include school names (as this could also be identify children).

Can we have a copy of the games? The games your child would play in this study are research software and are designed to help answer specific questions. They do not have the same type of functionality, amount of content, or style of documentation that you might expect from a commercial game. However, once the study is over the games will be made publicly available online, for free unlimited use. This should be in early 2015. You will be sent a link and starting instructions; all you will need is a computer with internet access.

Who paid for this research? This study is part of the PhD work for the main researcher, Alyssa Alcorn. It is indirectly paid for by University of Edinburgh-funded PhD scholarships, which are not attached to a specific project or to any outcomes of that project. Conducting this research brings no additional financial benefit to the researchers or to the university.

Would you like to participate?

We ask parents to read this information sheet and to decide whether they think it is a good idea for their child to participate as a games tester. Parents may also indicate their choices about how your child's information may used.

If you say "yes" when returning the permission form, this means that the researchers (or researchers and your child's teacher) will then explain the games tester role to your child, and ask him or her if s/he wants to help. We will remind the child that they can stop being a tester at any time, without having to give any reason, and that we will always listen to them. We will check that the child agrees to be video recorded. This explanation will be based on the child information sheet included in this packet. We feel strongly that children should be given a real choice about whether to participate--or not. Even if you say "yes" on the permission form, your child may still say "no" if s/he does not want to be a games tester. We will respect your child's decision. **If you say no,** then we will not contact you again about this study and will not speak to your child about being a games tester.

Once again, this study is *completely voluntary* and you and your child are under no obligation to take part. Even if you say yes now, you may withdraw your child from the study *at any time and for any reason* by contacting us, or your child's teacher. Your child may also withdraw at any time by saying that s/he does not want to be a games tester any more.

Thank you for taking the time to read this.

If you would like to know more about this research and/or if you have questions, please contact Ms. Alyssa Alcorn from the University of Edinburgh School of Informatics, (aalcorn@ed.ac.uk) or on *mobile number redacted*.

New computer games to help children
(to be read aloud to the child)

This page is for children. It is about some people who will visit your school soon. It says who they are, and what they will do at your school. The visitors will ask for your help. You can help by playing a new computer game. You can decide if you want to say “yes” or “no” to helping.

Who is visiting my school?

This is Alyssa and Helen. Their job is to learn about how children use computers, and how to make computer games that can help children. They will ask you to help them by playing a new computer game.

Alyssa

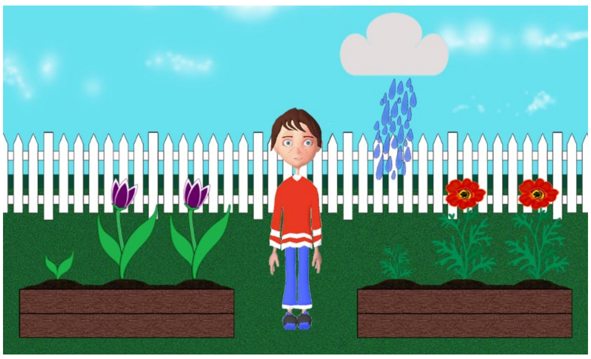
Helen

Why are Alyssa and Helen visiting my school?

It is hard for some children to talk to other people. Alyssa and Helen have a question about this. “Can playing special computer games help these children to talk to other people more?”

Alyssa and Helen will ask children like you to help them find the answer. You can help by being a “games tester” and playing a new computer game.

This game is about a Magic Garden. There are lots of things to do in the garden, but Alyssa and Helen do not know which things children will like best. Here is a picture of the Magic Garden:



Alyssa and Helen will ask children to play this game. This game is just for fun. It does not have any levels or points. No one wins or loses. Some children may think it is easy. Other children may think it is hard. That is OK. All children can help. Alyssa and Helen will learn when they see children play the game.

This page says how you can help as a games tester. Then you can decide if you want to say “yes” or “no” to helping.

What will happen if I help?

Alyssa and Helen will bring the game to your school. When your teacher says it is OK, you will go to [NAME OF ROOM]. There, you will get to play the game on a big touch screen. You will get to play the game on more than one day.

You can tell Alyssa and Helen if you want to stop playing the game. You do not have to tell them why. Please tell them if you need the toilet, or if you want to take a break. You can also say you do not want to be a games tester any more, and that is OK. Alyssa and Helen will always listen to you.

Alyssa and Helen will bring a video camera to your school. They will ask if it is OK to make a video of you playing the game. This is because it is too hard for Alyssa and Helen to write down everything that happens. They will watch the video later to help them understand what you said, and what happened when you played the game.

What will happen after I am finished helping?

Alyssa and Helen will learn a lot about children and computers when they watch you play the game. After all the children have finished playing the game, Alyssa and Helen will take the game and the touch screen away.

They will watch the videos of children playing the game. They will write about what they have learned. When Alyssa and Helen talk to other people who make games, they will say what they learned. Sometimes Alyssa and Helen will show people videos of children playing or talking about the game. Those things are easy to show in a video, but hard to say in words!

Later, Alyssa and Helen will send another page like this one to you and your parents. They will say “Thank you!” for your help. The page will say what you did to help them. It will explain what they learned about children and computers.

Your mum or dad said it is OK for you to help us. Your teacher said it is OK for you to help us.

Do you want to be a games tester? You can say “yes” or “no”. It is OK to say “no”. It will not hurt Alyssa and Helen's feelings.

Do you want to ask a question about being a games tester?

It is OK to have more questions. You can ask Alyssa as many questions as you want about being a games tester. Ask your mum or dad to help you call her on the phone or write an e-mail with your question.

Mobile phone number: *mobile number redacted*
E-mail address: aalcorn@ed.ac.uk

Research permission form (for parents)

	Please circle
Have you read the information sheet(s)?	YES / NO
Have you received enough information about the study?	YES / NO
Do you understand that participation is completely voluntary and your child can leave the study (or any individual activity) at any time, without having to give a reason?	YES / NO
Do you consent to completing the Social Communication Questionnaire (SCQ) form [1]?	YES / NO

Please sign this page to indicate that you understand and accept the conditions of this study, including video recording and background data collection through the British Picture Vocabulary Scale (BPVS). By signing, you agree that the researchers may explain the study to your child and invite him or her to take part as a games tester.

With reference to further anonymous use of photographic or video data, please circle yes or no in response to the following:

> I AGREE that short videos/images of my child can be used as examples in documents and presentations for research and/or teaching purposes.
> YES / NO

If you give permission for this study, please return this form and the completed SCQ form to your child's teacher.

If you DO NOT wish to give permission, you do not need to return this page. We will not ask your child to participate. However, please return the blank SCQ form.

Full name of participating child: ______________________________

Child's date of birth (DD/MM/YYYY): _______/_______/_______

Your relationship to the child: ______________________________

Your name (please print clearly):______________________________

Contact telephone number: ______________________________

e-mail address: ______________________________

Best time and method to reach you? ______________________________

Signature: ______________________________

Date: _______ / _______ /_______

[1] Only relevant if your child has a prior autism spectrum diagnosis

Appendix I

Supplementary documents for Study 2

Staff instructions were distributed to the school staff members who would accompany children to games study session, to inform them of the purpose of the study, and outline situations in which the researcher would—and would not—like them to intervene to support the child. The emphasis was on supporting the child should s/he become anxious or upset, rather than supporting the child to stay on-task or do the games correctly.

Games session goals: play and discovery

This study is about children playing games—it is not an assessment.

It is important to see what the children are interested in and what they can do independently. *It is much less important that they finish things or "do it right".*

It is OK if children:

- Are more interested in exploring individual features of the games than in finishing game tasks.
- Play very slowly or use repetitive behaviours.
- Stop to talk about the game or ask questions. Talking about the games is just as important for our research as playing the games!

How you can help during games sessions:

Please allow us to be the main people giving instructions, helping, and interacting with the child. *If we cannot help a child or explain something to him/her, then we may ask for your assistance.*

Please respond if a child wants to talk to you about the game, or show you things in the game. This is really positive!

If a child is struggling, please resist helping. We prefer children to be the ones to ask for help. If they do not, we will try to help them first.

Please speak up and step in if:

- The child is becoming very anxious or upset (or over-excited)
- The child needs the toilet or needs a break but has not said so
- The child's session should end for some other reason.

Finally, **it is very important that you do not point out events in the game that you think the child may have missed.** This is because a main purpose of the study is to find out what children spontaneously notice and react to during the game.

Thank you very much for your help and support!

References

3deverart (2010). *"cartoon boy andy" 3-D character model for computer animation*. http://www.turbosquid.com.

Adams, D. and McLaren, B. (2013). Erroneous Examples as Desirable Difficulty. In Lane, H. C., Yacef, K., Mostow, J., and Pavlik, P., editors, *16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013.*, pages 803–806.

Adobe Systems Incorporated (2010). *Adobe Photoshop Elements*. Adobe Systems Incorporated, 9.0.3 edition.

Alcorn, A. (2010). *Exploring Joint Attention Responses in Autism Spectrum Children: Interacting with a Virtual Agent in the ECHOES TEL Environment*. Unpublished master's thesis, University of Edinburgh, UK, Edinburgh, UK.

Alcorn, A., Good, J., and Pain, H. (2013a). Deliberate system-side errors as a potential pedagogic strategy for exploratory virtual learning environments. In Lane, H. C., Yacef, K., Mostow, J., and Pavlik, P., editors, *16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013.*, pages 483–492. Springer-Verlag Berlin Heidelberg.

Alcorn, A., Pain, H., and Good, J. (2013b). Discrepancies in a virtual learning environment: something worth communicating about for young children with ASC? In *Proceedings of the 12th International Conference on Interaction Design and Children*, pages 56–65. ACM.

Alcorn, A., Pain, H., Good, J., and Rajendran, G. (2012). Surprising Events in a Virtual Environment: A Catalyst for the Initiation of Spontaneous Social Interactions by Children with ASC. Poster presentation and demonstration. In *International Meeting for Autism Research technology demonstration session, Toronto, CA.*, Toronto, Canada. International Meeting for Autism Research.

Alcorn, A., Pain, H., Rajendran, G., Smith, T., Lemon, O., Porayska-Pomsta, K., Foster, M. E., Avramides, K., Frauenberger, C., and Bernardini, S. (2011). Social communication between virtual characters and children with autism. In Biswas, G., Bull, S., Kay, J., and Mitrovic, A., editors, *Artificial Intelligence in Education*, volume 6738, pages 7–14, Berlin. Springer Verlag.

Alcorn, A. M., Pain, H., and Good, J. (2014). Motivating children's initiations with novelty and surprise: initial design recommendations for autism. In *Proceed-*

ings of the 2014 conference on Interaction design and children, pages 225–228, Aarhus, Denmark. ACM.

Aleven, V., Stahl, E., Schworm, S., Fischer, F., and Wallace, R. (2003). Help Seeking and Help Design in Interactive Learning Environments. *Review of Educational Research*, 73(3):277–320.

Angen, M. J. (2000). Evaluating interpretive inquiry: Reviewing the validity debate and opening the dialogue. *Qualitative Health Research*, 10(3):378–395.

Anthony, L., Brown, Q., Nias, J., and Tate, B. (2013). Examining the need for visual feedback during gesture interaction on mobile touchscreen devices for kids. In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, pages 157–164, New York, New York, USA. ACM Press.

APA (2013). *Diagnostic and statistical manual of mental disorders : DSM-5*. American Psychiatric Association, Arlington, VA, 5th edition.

Ash, R., Crook, J., Dannenberg, R., Drung, B., Johnson, V., Lucius, L., Shaw, M., and Al., E. (2014). *Audacity® version 2.0.6*. Open source software project.

Bakeman, R. and Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child development*, 55(4):1278–89.

Baron-Cohen, S., Golan, O., and Ashwin, E. (2009). Can emotion recognition be taught to children with autism spectrum conditions? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364(1535):3567–74.

Benton, L. and Johnson, H. (2013). Structured approaches to participatory design for children: can targeting the needs of children with autism provide benefits for a broader child population? *Instructional Science*.

Berlyne, D. E. (1960). *Conflict, Arousal, and Curiosity*. McGraw Hill Book Company, Inc., New York, New York, USA.

Bernardini, S. and Porayska-Pomsta, K. (2013). Planning-Based Social Partners for Children with Autism. In *Proceedings of the Twenty-Third International Conference on Automated Planning and Scheduling Planning-Based*, pages 362–370. AAAI.

Boden, M. (1995). Creativity and Unpredictability. *Stanford Humanities Review*, 4(2):1–18.

Boden, M. (2010). *Creativity and Art: Three Roads to Surprise*. Oxford University Press, Oxford, UK.

Boden, M. A. (1990). *The Creative Mind: Myths and Mechanisms*. Abacus, London, 1992 edition.

Bono, M. a., Daley, L. T., and Sigman, M. (2004). Relations among joint attention, amount of intervention and language gain in autism. *Journal of autism and developmental disorders*, 34(5):495–505.

Breen, C. G. and Haring, T. G. (1991). Effects of contextual competence on social initiations. *Journal of Applied Behavior Analysis*, 24:337–347.

Brugman, H. and Russel, A. (2004). Annotating multi-media / multi-modal resources with ELAN. *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC 2004)*, (October):2065–2068.

Bruner, J. S. (1962). The conditions of creativity. In Gruber, H. E., Terrell, G., and Wertheimer, M., editors, *Contemporary Approaches to Creative Thinking*, chapter 1, pages 1–30. Atherton Press, New York, New York, USA.

Bruner, J. S. (1976). Surprise, craft, and creativity. In Bruner, J. S., Jolly, A., & Sylva, K., editor, *Play: Its role in development and evolution*, chapter 66, pages 641–642. Penguin, Harmondsworth.

Bryman, A. (1988). *Quantity and Quality in Social Research*. Unwin Hyman Ltd., London.

Butterworth, G. and Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, 9(1):55–72.

Byrt, T., Bishop, J., and Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, 46(5):423–429.

Carter, A. S., Messinger, D. S., Stone, W. L., Celimli, S., Nahmias, A. S., and Yoder, P. (2011). A randomized controlled trial of Hanen's 'More Than Words' in toddlers with early autism symptoms. *Journal of child psychology and psychiatry, and allied disciplines*, 52(7):741–52.

Carter, M. (2002). Communicative Spontaneity in Individuals with High Support Needs: An exploratory consideration of causation. *International Journal of Disability, Development and Education*, 49(3):225–242.

Charlesworth, W. R. (1969). Surprise and cognitive development. In Elkind, D. and Flavell, J. H., editors, *Studies in cognitive development: Essays in honor of Jean Piaget*, pages 257–314. Oxford University Press, New York, New York, USA.

Charlop, M., Kurtz, P., and Milstein, J. (1992). Too much reinforcement, too little behavior: assessing task interspersal procedures in conjunction with different reinforcement schedules with autistic children. *Journal of Applied Behavior Analysis*, 25(4):795.

Charlop, M. H. and Milstein, J. P. (1989). Teaching autistic children conversational speech using video modelling. *Journal of Applied Behaviour Analysis*, 3(22):275–285.

Charlop, M. H., Schreibman, L., and Tryon, A. S. (1983). Learning through observation: the effects of peer modeling on acquisition and generalization in autistic children. *Journal of abnormal child psychology*, 11(3):355–66.

Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philosophical Transactions: Biological Sciences*, 358:315–324.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Drew, A., and Cox, A. (2003). Predicting language outcome in infants with autism and pervasive developmental disorder. *International journal of language & communication disorders / Royal College of Speech & Language Therapists*, 38(3):265–85.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis.* SAGE Publications, Thousand Oaks, CA.

Chiang, H.-M. (2009). Differences between spontaneous and elicited expressive communication in children with autism. *Research in Autism Spectrum Disorders*, 3(1):214–222.

Chiang, H.-M. and Carter, M. (2008). Spontaneity of communication in individuals with autism. *Journal of autism and developmental disorders*, 38(4):693–705.

Constable, C. (1983). Creating communicative context. In Winitz, H., editor, *Treating language disorders: for clinicians by clinicians*. University Park Press, Baltimore, MD.

Constantin, A. (2015). *Supporting Practitioners in Social Story Interventions: The ISISS Authoring Tool*. PhD thesis, University of Edinburgh.

Cronbach, L. J. (1975). Beyond the Two Disciplines of Scientific Psychology. *American Psychologist*, 30(2):116–127.

Curcio, F. and Paccia, J. (1987). Conversations with autistic children: contingent relationships between features of adult input and children's response adequacy. *Journal of autism and developmental disorders*, 17(1):81–93.

Dautenhahn, K., Werry, I., Robins, B., Dickerson, P., and Stribling, P. (2004). Robot-mediated joint attention in children with autism: A case study in robot-human interaction. *Interaction Studies*, 5(2):161–198.

Davidson, D. (2001). The emergence of thought. In *Subjecting, Intersubjective, Objective*, pages 123–134. Clarendon Press, Oxford, UK.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., and Liaw, J. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental psychology*, 40(2):271–83.

Department of Health (2009). Valuing People Now: a new three-year strategy for people with learning disabilities. Technical report, HM Government.

Dettmer, S., Simpson, R. L., Myles, B. S., and Ganz, J. B. (2000). The Use of Visual Supports to Facilitate Transitions of Students with Autism. *Focus on Autism and Other Developmental Disabilities*, 15(3):163–169.

Dey, I. (1993). *Qualitative data analysis: a user-friendly guide for social scientists*. Routledge, London.

Druin, A. (2002). The role of children in the design of new technology. *Behaviour and Information Technology*, 21(1):1–25.

DSM-IV, A. P. A. T. F. (1994). *DSM-IV: diagnostic and statistical manual of mental disorders*. American Psychiatric Association Washington, DC.

Dunn, L. M., Dunn, D. M., Whetton, C. W., and Burley, J. (1997). *The British Picture Vocabulary Scale*. NFER Nelson, Windsor, UK, second edition.

Dunn, L. M., Dunn, D. M., &. N. (2009). *The British Picture Vocabulary Scale*. GL Assessment Ltd, London, 3rd edition.

Eigsti, I. M., De Marchena, A. B., Schuh, J. M., and Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, 5(2):681–691.

Ferrara, C. and Hill, S. D. (1980). The responsiveness of autistic children to the predictability of social and nonsocial toys. *Journal of Autism and Developmental Disorders*, 10(1):51–57.

Fisher, W. W., Pawich, T. L., Dickes, N., Paden, A. R., and Toussaint, K. (2014). Increasing the saliency of behavior-consequence relations for children with autism who exhibit persistent errors. *Journal of Applied Behavior Analysis*, 47(4):738–748.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitivedevelopmental inquiry. *American Psychologist*, 34(10):906–911.

Fletcher-Watson, S. (2013). A Targeted Review of Computer-Assisted Learning for People with Autism Spectrum Disorder: Towards a Consistent Methodology. *Review Journal of Autism and Developmental Disorders*, 1(2):87–100.

Fletcher-Watson, S. and Durkin, K. (2015). Uses of new technologies by young people with developmental disorders. In Van Herwegen, J., & Riby, D., editor, *Neurodevelopmental disorders: research challenges and solutions*, chapter 13, pages 243–267. Pscyhology Press, London.

Fletcher-Watson, S., Petrou, A., Scott-Barrett, J., Dicks, P., Graham, C., O'Hare, A., Pain, H., and McConachie, H. (2015). A trial of an iPad intervention targeting social communication skills in children with autism. *Autism : the international journal of research and practice*, pages 1–12.

Foster, M., Avramides, K., Bernardini, S., Chen, J., Frauenberger, C., Lemon, O., and Porayska-Pomsta, K. (2010). Supporting Children's Social Communication Skills through Interactive Narratives with Virtual Characters. *Proceedings of the international conference on Multimedia*, (Figure 2):1111–1114.

Fraser, S., Lewis, V., Ding, S., Kellett, M., and Robinson, C., editors (2004). *Doing Research with Children and Young People*. SAGE Publications, London.

Frauenberger, C., Good, J., Alcorn, A., and Pain, H. (2012a). Supporting the design contributions of children with autism spectrum conditions. *Proceedings of the 11th International Conference on Interaction Design and Children - IDC '12*, page 134.

Frauenberger, C., Good, J., Alcorn, A., and Pain, H. (2013). Conversing through and about technologies: Design critique as an opportunity to engage children with autism and broaden research(er) perspectives. *International Journal of Child-Computer Interaction*, 1(2):38–49.

Frauenberger, C., Good, J., and Keay-Bright, W. (2011). Designing technology for children with special needs: bridging perspectives through participatory design. *CoDesign*, 7(1):1–28.

Frauenberger, C., Good, J., Keay-Bright, W., and Pain, H. (2012b). Interpreting input from children: a designerly approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2377–2386, Austin, TX. ACM.

Gall, S. N. L. (1981). Help-seeking: An understudied problem-solving skill in children. *Developmental Review*, 1(3):224–246.

GameSalad Inc. (2014). *GameSalad Creator beta, version 0.10.5*. GameSalad, Inc., Austin, TX, windows edition.

Gaver, W. (2014). Science and Design: The Implications of Different Forms of Accountability. In Olsen, J. S. and Kellogg, W. A., editors, *Ways of Knowing in HCI*, pages 143–165. Springer, New York.

Geertz, C. (1973). *The interpretation of cultures: Selected essays*. Basic Books, New York.

Glaser, B. and Strauss, A. (1967). *The discovery of grounded theory. 1967*. Aldine Publishing Company, New York, NY, USA.

Goldsmith, T. R. and Leblanc, L. A. (2004). Use of Technology in Interventions for Children with Autism. 1(2):166–178.

Goodhart, F. and Baron-Cohen, S. (1993). How many ways can the point be made? Evidence from children with and without autism. *First Language*, 13(38):225–233.

Gotsis, M., Piggot, J., Hughes, D., and Stone, W. (2010). SMART-games: a video game intervention for children with Autism Spectrum Disorders. In *Proceedings of the 9th International Conference on Interaction Design and Children*, pages 194–197. ACM.

Grawemeyer, B., Johnson, H., Brosnan, M., Ashwin, E., and Benton, L. (2012). Developing an embodied pedagogical agent with and for young people with autism spectrum disorder. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 7315 LNCS, pages 262–267.

Grynszpan, O., Weiss, P. L., Perez-Diaz, F., and Gal, E. (2014). Innovative technology-based interventions for autism spectrum disorders: A meta-analysis. *Autism*, 18(4):346–361.

Hacker, D. J. (1998). Definitions and Empirical Foundations. In Hacker, D. J., Dunlosky, J., and Graesser, A. C., editors, *Metacognition in Educational Theory and Practice*, chapter 1, pages 1–23. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Hammersley, M. (1992a). Some reflections on ethnography and validity. *International Journal of Qualitative Studies in Education (QSE)*, 5(3):195–203.

Hammersley, M. (1992b). *What's Wrong with Ethnography: Methodological Explorations*. Routledge, London.

Harjusola-Webb, S. M. and Robbins, S. H. (2011). The Effects of Teacher-Implemented Naturalistic Intervention on the Communication of Preschoolers With Autism. *Topics in Early Childhood Special Education*, 32(2):99–110.

Hausman, C. R. (1975). *A Discourse on Novelty and Creation*. Martinus Nijhoff, The Hague, Netherlands.

Hayes, G. R., Hirano, S., Marcu, G., Monibi, M., Nguyen, D. H., and Yeganyan, M. (2010). Interactive visual supports for children with autism. *Personal and Ubiquitous Computing*, 14(7):663–680.

Hood, B. M. (1995). Gravity rules for 2- to 4-year olds? *Cognitive Development*, 10(4):577–598.

Hood, B. M. (1998). Gravity does rule for falling events. *Developmental Science*, 1(1):59–63.

Hourcade, J. P., Bullock-Rest, N. E., and Hansen, T. E. (2011). Multitouch tablet applications and activities to enhance the social skills of children with autism spectrum disorders. *Personal and Ubiquitous Computing*, 16(2):157–168.

Hourcade, J. P., Williams, S. R., Miller, E. a., Huebner, K. E., and Liang, L. J. (2013). Evaluation of tablet apps to encourage social interaction in children with autism spectrum disorders. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, page 3197.

Howlin, P. (1998). *Children with autism and Asperger syndrome: a guide for practitioners and carers*. John Wiley, Chichester.

Howlin, P. (2006). Augmentative and alternative communication communication systems for children with autism. In Charman, T. and Stone, W., editors, *Social and communication development in autism spectrum disorders: Early identification, diagnosis, and intervention*, chapter 9, pages 236–266. Guilford Press, New York, NY.

Hughes, T. and Helling, M. K. (1991). A case for obtaining informed consent from young children. *Early Childhood Research Quarterly*, 6(2):225–232.

Hutchinson, N. and Bodicoat, A. (2015). The Effectiveness of Intensive Interaction, A Systematic Literature Review. *Journal of Applied Research in Intellectual Disabilities*, 28(6):437–454.

Inkpen, K. M. (2001). Drag-and-drop versus point-and-click mouse interaction styles for children. *ACM Transactions on Computer-Human Interaction*, 8(1):1–33.

Joiner, R., Messer, D., Light, P., and Littleton, K. (1998). It is best to point for young children: a comparison of children's pointing and dragging. *Computers in Human Behavior*, 14(3):513–529.

Jolliffe, T., Lansdown, R., and Robinson, C. (2001). Autism: A personal account. In Paechter, C., editor, *Learning, Space, and Identity*, pages 42–56. Paul Chapman Publishing, London.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous child*, 2(3):217–250.

Kaplan, A. (1964). *The conduct of inquiry: methodology for behavioral science*. Chandler Publishing, San Francisco.

Karsten, A. M. and Carr, J. E. (2009). The effects of differential reinforcement of unprompted responding on the skill acquisition of children with autism. *Journal of applied behavior analysis*, 42(2):327–334.

Kasari, C., Freeman, S., and Paparella, T. (2006). Joint attention and symbolic play in young children with autism: a randomized controlled intervention study. *Journal of child psychology and psychiatry, and allied disciplines*, 47(6):611–20.

Kasari, C., Gulsrud, A., Freeman, S., Paparella, T., and Hellemann, G. (2012). Longitudinal Follow Up of Children with Autism Receiving Targeted Interventions on Joint Attention and Play. *Journal of the American Academy of Child and Adolescent Psychiatry*, 51(5):487–495.

Kasari, C., Gulsrud, A. C., Wong, C., Kwon, S., and Locke, J. (2010). Randomized controlled caregiver mediated joint engagement intervention for toddlers with autism. *Journal of Autism and Developmental Disorders*, 40(9):1045–1056.

Keay-Bright, W. (2007). Can computers create relaxation? Designing ReacTickles software with children on the autistic spectrum. *CoDesign*, 3(2):97–110.

Keay-Bright, W. (2008). ReacTickles Global : A Non-Textual Mobile & Networked Play Space. *Journal of Assistive Technologies*, 2(3):261–264.

Keay-Bright, W. and Howarth, I. (2012). Is simplicity the key to engagement for children on the autism spectrum? *Personal and Ubiquitous Computing*, 16(2):129–141.

Kellett, M. (2000). Sam's story: evaluating Intensive Interaction in terms of its effect on the social and communicative ability of a young child with severe learning difficulties. *Support for Learning*, 15(4):165–171.

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., and Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism : the international journal of research and practice*, 20(4):442–62.

Knight, V., McKissick, B. R., and Saunders, A. (2013). A review of technology-based interventions to teach academic skills to students with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 43(11):2628–2648.

Kossyvaki, L., Jones, G., and Guldberg, K. (2012). The effect of adult interactive style on the spontaneous communication of young children with autism at school. *British Journal of Special Education*, 39(4):173–184.

Layton, T. L. and Watson, L. R. (1995). Enhancing communication in nonverbal children with autism. In Quill, K. A., editor, *Teaching children with autism: Strategies to enhance communication and socialization*, chapter 4, pages 73–101. Delamar, Cengage Learning, Albany, NY.

Leekam, S. R. and Ramsden, C. a. H. (2006). Dyadic orienting and joint attention in preschool children with autism. *Journal of autism and developmental disorders*, 36(2):185–97.

Levy, S. E., Mandell, D. S., and Schultz, R. T. (2009). Autism. *Lancet*, 374(9701):1627–38.

Lewis, J. and Ritchie, J. (2003). Generalising from Qualitative Research. In Jane Ritchie and Jane Lewis, editor, *Qualitative research practice : a guide for social science students and researchers*, chapter 10, pages 263–286. SAGE Publications, London.

Lincoln, Y. S. and Guba, E. G. (1985). *Naturalistic inquiry*. SAGE, Newbury Park, Calif.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., and Rutter, M. (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*, 30(3):205–223.

Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of autism and developmental disorders*, 24(5):659–85.

Lord, C., Wagner, A., Rogers, S., Szatmari, P., Aman, M., Charman, T., Dawson, G., Durand, V. M., Grossman, L., Guthrie, D., Harris, S., Kasari, C., Marcus, L., Murphy, S., Odom, S., Pickles, A., Scahill, L., Shaw, E., Siegel, B., Sigman, M., Stone, W., Smith, T., and Yoder, P. (2005). Challenges in evaluating psychosocial interventions for Autistic Spectrum Disorders. *Journal of autism and developmental disorders*, 35(6):695–708; discussion 709–11.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of consulting and clinical psychology*, 55(1):3–9.

Loveland, K. a. and Landry, S. H. (1986). Joint attention and language in autism and developmental language delay. *Journal of autism and developmental disorders*, 16(3):335–49.

Mahoney, G. and MacDonald, J. (2003). *Responsive teaching: Parent-mediated developmental intervention*. Case Western Reserve University, Cleveland, OH.

Malinverni, L., Mora-Guiard, J., Padillo, V., Valero, L., Hervás, A., and Pares, N. (2016). An inclusive design approach for developing video games for children with Autism Spectrum Disorder. *Computers in Human Behavior*, In press.

Malinverni, L., MoraGuiard, J., Padillo, V., Mairena, M., Hervás, A., and Pares, N. (2014). Participatory design strategies to enhance the creative contribution of children with special needs. In *Proceedings of the 2014 conference on Interaction design and children - IDC '14*, pages 85–94, New York, New York, USA. ACM Press.

Malone, T. and Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude, learning, and instruction*, 3:223–253.

Martin, P. and Turner, B. (1986). Grounded Theory and Organizational Research. *The Journal of Applied Behavioral Science*, 22(2):141–157.

Max Planck Institute for Psycholinguistics (2012). *ELAN Linguistics Annotator, version 4.4.0*. The Language Archive, Nijmegen, The Netherlands.

Max Planck Institute for Psycholinguistics (2015). *ELAN Linguistics Annotator, Version 4.9.1*. The Language Archive, Nijmegen, The Netherlands.

May, R. B. and Macpherson, D. F. (1971). Size discrimination in children facilitated by changes in task difficulty. *Journal of Comparative and Physiological Psychology*, 75(3):453–458.

McClenny, C. S., Roberts, J. E., and Layton, T. L. (1992). Unexpected events and their effect on children's language. *Child Language Teaching and Therapy*, 8(3):229–245.

McLaren, B., Adams, D., and Durkin, K. (2012). To err is human, to explain and correct is divine: a study of interactive erroneous examples with middle school math students. In Ravenscroft, A., Lindstaedt, S., Kloos, C. D., and Hernández-Leo, D., editors, *21st Century Learning for 21st Century Skills: 7th European Conference of Technology Enhanced Learning, EC-TEL 2012*, pages 222–235. Springer Berlin Heidelberg.

Millen, L., Cobb, S. V. G., Patel, H., and Glover, T. (2012). Collaborative virtual environment for conducting design sessions with students with autism spectrum conditions. *Proc. 9th Intl Conf. on Disability, Virtual Reality and Assoc. Technologies*, pages 10–12.

Mundy, P. and Crowson, M. (1997). Joint attention and early social communication: implications for research on intervention with autism. *Journal of autism and developmental disorders*, 27(6):653–76.

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., and Seibert, J. (2003). A Manual for the Abridged Early Social Communication Scales (ESCS).

Mundy, P. and Newell, L. (2007). Attention, Joint Attention, and Social Cognition. *Current directions in psychological science : a journal of the American Psychological Society*, 16(5):269–274.

Mundy, P., Sigman, M., and Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of autism and developmental disorders*, 20(1):115–28.

Nind, M. (1996). Efficacy of Intensive Interaction: developing sociability and communication in people with severe and complex learning difficulties using an approach based on caregiver-infant interaction. *European Journal of Special Needs Education*, 11(1):48–66.

Nind, M. and Hewett, D. (1994). *Access to Communication: Developing the Basics of communication in People with Severe Learning Difficulties through Intensive Interaction*. Fulton, London.

NRC (2001). *Educating Children with Autism*. National Academy Press, Washington, DC.

Parsons, S. and Cobb, S. (2011). State-of-the-art of virtual reality technologies for children on the autism spectrum. *European Journal of Special Needs Education*, 26(3):355–366.

Parsons, S., Guldberg, K., Macleod, A., Jones, G., Prunty, A., and Balfe, T. (2009a). International review of the literature of evidence of best practice provision in the education of persons with autistic spectrum disorders. Technical report, The National Council for Special Education (Ireland).

Parsons, S., Guldburg, K., MacLeod, A., Jones, G. E., Prunty, A., and Balfe, T. (2009b). International review of the literature of evidence of best practice provision in the education of persons with Autistic Spectrum Disorders. Technical report, National Council for Special Education: Ireland.

Parsons, S., Leonard, A., and Mitchell, P. (2006). Virtual environments for social skills training: comments from two adolescents with autistic spectrum disorder. *Computers & Education*, 47(2):186–206.

Parsons, S., Mitchell, P., and Leonard, A. (2004). The use and understanding of virtual environments by adolescents with autistic spectrum disorders. *Journal of autism and developmental disorders*, 34(4):449–66.

Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods*. SAGE, Thousand Oaks, CA.

Pennington, R. C. (2010). Computer-Assisted Instruction for Teaching Academic Skills to Students With Autism Spectrum Disorders: A Review of Literature. *Focus on Autism and Other Developmental Disabilities*, 25(4):239–248.

Pickard, K. E. and Ingersoll, B. R. (2015). Brief Report: High and Low Level Initiations of Joint Attention, and Response to Joint Attention: Differential Relationships with Language and Imitation. *Journal of Autism and Developmental Disorders*, 45:262–268.

Pidgeon, N. and Henwood, K. (2004). Grounded theory. In Hardy, M. and Bryman, A., editors, *Handbook of Data Analysis*, chapter 28, pages 625–648. SAGE Publications Ltd, London.

Ploog, B. O., Scharf, A., Nelson, D., and Brooks, P. J. (2012). Use of Computer-Assisted Technologies (CAT) to Enhance Social, Communicative, and Language Development in Children with Autism Spectrum Disorders. *Journal of autism and developmental disorders*.

Porayska-Pomsta, K., Frauenberger, C., Pain, H., Rajendran, G., Smith, T., Menzies, R., Foster, M. E., Alcorn, A., Wass, S., Bernadini, S., Avramides, K., Keay-Bright, W., Chen, J., Waller, A., Guldberg, K., Good, J., and Lemon, O. (2012). Developing technology for autism: an interdisciplinary approach. *Personal and Ubiquitous Computing*, 16(2):117–127.

Prizant, B. and Rydell, P. (1984). Analysis of functions of delayed echolalia in autistic children. *Journal of Speech, Language and Hearing Research*, 27(June):183–192.

Prizant, B., Wetherby, A., Rubin, E., and Laurent, A. (2003). The SCERTS model: A transactional, family-centered approach to enhancing communication and socioemotional abilities of children with autism spectrum disorder. *Infants & Young Children*, 16(4):296.

Prizant, B., Wetherby, A., Rubin, E., Laurent, A., and Rydell, P. (2005). *The SCERTS [TM] Model: A Comprehensive Educational Approach for Children with Autism Spectrum Disorders.* Brookes Publishing Company, Baltimore, MD.

Prizant, B., Wetherby, A., Rubin, E., Laurent, A., and Rydell, P. (2006). *The SCERTS Model: A Comprehensive Educational Approach for Children with Autism Spectrum Disorders. (Volume I: Assessment).* Paul H. Brookes, Baltimore, MD.

Prizant, B. M. and Duchan, J. F. (1981). The functions of immediate echolalia in autistic children. *Journal of Speech & Hearing Disorders*, 46(3):211–249.

Putnam, C. and Chong, L. (2008). Software and technologies designed for people with autism: what do users want? In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, pages 3–10. ACM.

Quill, K. A. (1995a). Enhancing children's social-communicative interactions. In Quill, K. A., editor, *Teaching children with autism: Strategies to enhance communication and socialization*, chapter 7, pages 163–190. Delamar, Cengage Learning, Albany, NY.

Quill, K. A., editor (1995b). *Teaching Children With Autism: Strategies to Enhance Communication and Socialization*. Delamar, Cengage Learning, Albany, NY.

Rao, P. A., Beidel, D. C., and Murray, M. J. (2008). Social skills interventions for children with Asperger's syndrome or high-functioning autism: a review and recommendations. *Journal of autism and developmental disorders*, 38(2):353–61.

Ritchie, J. (2003). The Application of Qualitative Methods to Social Research. In Ritchie, J. and Lewis, J., editors, *Qualitative research practice : a guide for social science students and researchers*, chapter 2, pages 24–46. SAGE Publications Ltd, London.

Robins, B. and Dautenhahn, K. (2006). The role of the experimenter in HRI research–a case study evaluation of children with autism interacting with a robotic toy. *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2006)*, 44(0):646–651.

Rogers, S. (2006). Evidence-based interventions for language development in young children with autism. In Charman, T. and Stone, W., editors, *Social and communication development in autism spectrum disorders: Early identification, diagnosis, and intervention*, chapter 6, pages 143–179. The Guilford Press, New York, NY.

Russell, J., Jarrold, C., and Hood, B. (1999). Two intact executive capacities in children with autism: implications for the core executive dysfunctions in the disorder. *Journal of autism and developmental disorders*, 29(2):103–12.

Rutter, M., Le Couteur, A., and Lord, C. (2003). *Autism diagnostic interview, revised*. Western Psychological Services, Los Angeles, CA.

Saldaña, J. (2009). *The Coding Manual for Qualitative Researchers*. SAGE Publications, London.

Schmidt, C. and Schmidt, M. (2008). Three-dimensional virtual learning environments for mediating social skills acquisition among individuals with autism spectrum disorders. In *Proceedings of the 7th international conference on Interaction design and children*, pages 85–88. ACM.

Schreibman, L. (2005). Characteristics of Autism. In *The Science and Fiction of Autism*, chapter 2, pages 26–48. Harvard University Press, Cambridge, MA.

Schuler, A. L. (1979). Echolalia: Issues and Clinical Applications. *Journal of Speech and Hearing Disorders*, 44(4):41–434.

Seale, C. (1999). Quality in Qualitative Research. *Qualitative Inquiry*, 5(4):465–478.

Seibert, J. M., Hogan, A. E., and Mundy, P. C. (1982). Assessing interactional competencies: The early social-communication scales. *Infant Mental Health Journal*, 3(4):244–258.

Shane, H. C. and Albert, P. D. (2008). Electronic screen media for persons with autism spectrum disorders: results of a survey. *Journal of autism and developmental disorders*, 38(8):1499–508.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Infortmation*, 22:63–75.

Sigman, M. and McGovern, C. W. (2005). Improvement in cognitive and language skills from preschool to adolescence in autism. *Journal of autism and developmental disorders*, 35(1):15–23.

Sigman, M., Ruskin, E., Mervis, C. B., and Robinson, B. F. (1999). *Continuity and change in the social competence of children with autism, Down syndrome, and developmental delays*. Wiley-Blackwell.

Silverman, D. (2006). *Interpreting qualitative data: methods for analysing talk, text and interaction*. SAGE, London, third edition.

Silverman, D. (2011). *Interpreting qualitative data: A guide to the principles of qualitative research*. SAGE, London, 4th edition.

Smith, T., Scahill, L., Dawson, G., Guthrie, D., Lord, C., Odom, S., Rogers, S., and Wagner, A. (2007). Designing research studies on psychosocial interventions in autism. *Journal of autism and developmental disorders*, 37(2):354–66.

Snape, D. and Spencer, L. (2003). The Foundations of Qualitative Research. In Ritchie, J. and Lewis, J., editors, *Qualitative research practice : a guide for social science students and researchers*, chapter 1, pages 1–23. SAGE Publications Ltd, London.

Snyder-McLean, L., Solomonson, B., McLean, J., and Sack, S. (1984). Structuring joint action routines: A strategy for facilitating communication and language development in the classroom. *Seminars in Speech and Language*, 5:213–228.

Stake, R. (1978). The case study method in social inquiry. *Educational Researcher*, 7(2):5–8.

Stanley, H. M. (1895). *Studies in the evolutionary psychology of feeling*. Swan Sonnenschein & Co., London.

Stone, W. L. and Caro-Martinez, L. M. (1990). Naturalistic observations of spontaneous communication in autistic children. *Journal of autism and developmental disorders*, 20(4):437–53.

Strain, P. S. (1977). An experimental analysis of peer social initiations on the behavior of withdrawn preschool children: Some training and generalization effects. *Journal of Abnormal Child Psychology*, 10(2):289–298.

Strauss, A. and Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. SAGE Publications, Thousand Oaks, CA, 2nd edition.

Sussman, F. (1999). *More than words: Helping parents promote communication and social skills in children with autism spectrum disorders*. The Hanen Centre, Toronto, Canada.

Tager-Flusberg, H. and Anderson, M. (1991). The development of contingent discourse ability in autistic children. *Journal of child psychology and psychiatry, and allied disciplines*, 32(7):1123–34.

Tanaka, J. W., Wolf, J. M., Klaiman, C., Koenig, K., Cockburn, J., Herlihy, L., Brown, C., Stahl, S., Kaiser, M. D., and Schultz, R. T. (2010). Using computerized games to teach face recognition skills to children with autism spectrum disorder: the Let's Face It! program. *Journal of child psychology and psychiatry, and allied disciplines*, 51(8):944–52.

Tartaro, A. and Cassell, J. (2008). Playing with virtual peers: bootstrapping contingent discourse in children with autism. In *Proceedings of the 8th international conference on International conference for the learning sciences-Volume 2*, pages 382–389. International Society of the Learning Sciences.

The Scottish Government (2013). Menu of Interventions. Technical report, The Scottish Government, Edinburgh.

Tomasello, M. (1995). Joint Attention as Social Cognition. In Moore, C. and Dunham, P. J., editors, *Joint Attention: Its Origins and Role in Development*, chapter 6, pages 103–130. Taylor & Francis, Hillsdale, NJ.

Tsovaltzi, D., McLaren, B., Melis, E., and Meyer, A.-K. (2012). Erroneous examples: effects on learning fractions in a web-based setting. *International Journal of Technology Enhanced Learning*, 4(3/4):191–227.

Turner, B. (1981). Some Practical Aspects of Qualitative Data Analysis: One Way Of Organising the Cognative Processes Associated with the Generation of Grounded Theory. *Quality & Quantity*, 15(4):225–248.

Twatchman, D. D. (1995). Methods to enhance communication in verbal children. In Quill, K. A., editor, *Teaching children with autism: Strategies to enhance communication and socialization*, chapter 6, pages 133–162. Delamar, Cengage Learning, Albany, NY.

van Rijn, H. and Stappers, P. J. (2008). The Puzzling Life of Autistic Toddlers: Design Guidelines from the LINKX Project. *Advances in Human-Computer Interaction*, 2008:1–8.

Venter, A., Lord, C., and Schopler, E. (1992). A Follow-Up Study of High-Functioning Autistic Children. *Journal of Child Psychology and Psychiatry*, 33(3):489–597.

Vismara, L. A. and Lyons, G. L. (2007). Using Perseverative Interests to Elicit Joint Attention Behaviors in Young Children With Autism: Theoretical and Clinical Implications for Understanding Motivation. *Journal of Positive Behavior Interventions*, 9(4):214–228.

Wainer, A. L. and Ingersoll, B. R. (2011). The use of innovative computer technology for teaching social communication to individuals with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 5(1):96–107.

Wainer, J., Robins, B., Amirabdollahian, F., and Dautenhahn, K. (2014). Using the humanoid robot KASPAR to autonomously play triadic games and facilitate collaborative play among children with autism. *IEEE Transactions on Autonomous Mental Development*.

Walden, T. and Hurley, J. (2006). A developmental approach to understanding atypical development. In Charman, T. and Stone, W., editors, *Social and communication development in autism spectrum disorders: early identification, diagnosis, and intervention*, chapter 10, pages 269–295. The Guilford Press, New York, NY.

Wall, K. (2010). *Autism and Early Years Practice*. SAGE Publications Ltd, London, 2nd edition.

Wetherby, A. (2006). Understanding and measuring social communication in children with autism spectrum disorders. In Charman, T. and Stone, W., editors, *Social and communication development in autism spectrum disorders: Early identification, diagnosis, and intervention*, chapter 1, pages 3–34. The Guilford Press, New York, NY.

Wetherby, A. and Prizant, B. (1989). The expression of communicative intent: Assessment guidelines. *Seminars in Speech and Language*, 10(1):77–91.

Wetherby, A. and Prutting, C. (1984). Profiles of communicative and cognitive-social abilities in autistic children. *Journal of speech and hearing research*, 27(3):364–377.

Whalen, C., Liden, L., and Ingersoll, B. (2006a). Behavioral improvements associated with computer-assisted instruction for children with developmental disabilities. *The Journal of*, 1(1):11–26.

Whalen, C., Moss, D., Ilan, A. B., Vaupel, M., Fielding, P., Macdonald, K., Cernich, S., and Symon, J. (2010). Efficacy of TeachTown: Basics computer-assisted intervention for the Intensive Comprehensive Autism Program in Los Angeles Unified School District. *Autism : the international journal of research and practice*, 14(3):179–197.

Whalen, C. and Schreibman, L. (2003). Joint attention training for children with autism using behavior modification procedures. *Journal of Child Psychology and Psychiatry*, 44(3):456–468.

Whalen, C., Schreibman, L., and Ingersoll, B. (2006b). The collateral effects of joint attention training on social initiations, positive affect, imitation, and spontaneous speech for young children with autism. *Journal of autism and developmental disorders*, 36(5):655–664.

White, P. J., O'Reilly, M., Streusand, W., Levine, A., Sigafoos, J., Lancioni, G., Fragale, C., Pierce, N., and Aguilar, J. (2011). Best practices for teaching joint attention: A systematic review of the intervention literature. *Research in Autism Spectrum Disorders*, 5(4):1283–1295.

Wieder, S. and Greenspan, S. I. (2003). Climbing the symbolic ladder in the DIR model through floor time/interactive play. *Autism : the international journal of research and practice*, 7(4):425–435.

Wing, L. A. (1995). Play is not the work of the child: Young children's perceptions of work and play. *Early Childhood Research Quarterly*, 10:223–247.

Woolf, B. P. (2009). *Building Intelligent Interactive Tutors: Student-centered strategies for revolutionizing e-learning*. Morgan Kaufmann Publishers, Burlington, Massachusetts.

Yeganyan, M. T., Cramer, M., Boyd, L. A., and Hayes, G. R. (2010). vSked: an interactive visual schedule system for use in classrooms for children with autism. In *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10*, pages 319–322.

Yoder, P. and McDuffie, A. S. (2006). Treatment of Responding to and Intitiating Joint Attention. In Charman, T. and Stone, W., editors, *Social and communication development in autism spectrum disorders: Early identification, diagnosis, and intervention*, chapter 5, pages 117–142. The Guilford Press, New York, New York, USA.